

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE  
UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À  
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

COMME EXIGENCE PARTIELLE  
À L'OBTENTION DE LA  
MAÎTRISE EN GÉNIE DE LA CONSTRUCTION  
M.ING

PAR  
HAKIM TAHRI

FILTRAGE ET ESTIMATION UNIVARIÉS DES DÉBITS ET DES  
CONCENTRATIONS DE POLLUANTS

MONTREAL /SEPTEMBRE 2001

© droits réservés de Hakim Tahri 2001

**CE MÉMOIRE A ÉTÉ ÉVALUÉ**

**PAR UN JURY COMPOSÉ DE :**

- M. Saad Bennis, directeur de mémoire  
Département de génie de la construction à l'École de technologie supérieure
- M. Robert Leconte, professeur  
Département de génie de la construction à l'École de technologie supérieure
- M. Joseph Michelet, ingénieur  
Communauté urbaine de Montréal

**IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC**

**LE 5 AVRIL 2001**

**À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE**

# **FILTRAGE ET ESTIMATION UNIVARIÉS DES DÉBITS ET DES CONCENTRATIONS DE POLLUANTS**

Hakim Tahri

## **Sommaire**

En temps de pluie, le citoyen se voit menacé d'inondations qui peuvent provenir aussi bien d'insuffisances dans les réseaux d'assainissement que des débordements des cours d'eau. Cette augmentation des débits provoque aussi un dépassement de la capacité des stations d'épuration et le déversement des surplus bruts d'eaux pluviales mélangées avec des eaux usées dans des rivières, leur faisant ainsi perdre la plupart de leurs usages.

La gestion en temps réel des réseaux d'assainissement, notamment unitaires, est considérée comme un moyen efficace d'optimiser les installations existantes dans le cadre d'une protection accrue du citoyen et de l'environnement. L'amélioration des technologies de l'information est à la base d'un renouveau dans cette gestion dynamique des réseaux. La centralisation des données recueillies en temps réel sur la qualité et la quantité des flux à différents endroits du réseau permet d'agir à distance pour aiguiller ces flux et utiliser d'une manière optimale la capacité de transport et de stockage afin de minimiser l'impact des déversements et des inondations. Malheureusement, les données recueillies sont souvent erronées et parfois manquantes. Le filtrage des mesures constitue un élément clé dans la réussite d'un bon diagnostic pour évaluer sur une longue période la performance du système d'assainissement vis-à-vis des contraintes réglementaires. Ce travail pour objectif de filtrer les données relatives à la quantité et la qualité du ruissellement en milieu urbain.

Les données filtrées peuvent aussi bien être utilisées en temps différé pour concevoir de nouveaux ouvrages d'assainissement, évaluer des ouvrages existants en vue de les maintenir et de les réhabiliter qu'en temps réel pour assurer une gestion optimale des ouvrages orientée vers l'amélioration de la qualité du milieu récepteur et la sécurité civile.

## **REMERCIEMENTS**

En témoignage de ma reconnaissance je tiens à exprimer mes vifs remerciements à Monsieur Saad Bennis mon Directeur de projet et professeur au département de la construction à l'ÉTS pour son soutien moral, ces conseils pertinents et pour l'intérêt avec lequel il a suivi la progression de mon travail.

Je tiens aussi à remercier Monsieur Faouzi Berrada professeur à la faculté des sciences Ain Chock, université de Hassan II, Casablanca, Maroc pour la transmission de ses expériences, ainsi que son soutien moral.

Je tiens aussi à faire-part de mes remerciements à tous ceux qui ont participé de proche ou de loin à ce travail.

Finalement je dédis ce travail à ma mère Khadija, ainsi que ma sœur Ghita qui m'ont procuré un soutien exceptionnel le long de mon cursus universitaire.

# TABLE DES MATIÈRES

	Page
SOMMAIRE.....	i
REMERCIEMENTS.....	ii
LISTE DES TABLEAUX.....	vi
LISTE DES FIGURES.....	vii
LISTE DES ABRÉVIATIONS ET SIGLES.....	ix
INTRODUCTION.....	1
CHAPITRE 1 : Contexte général du projet.....	3
1.1 Introduction.....	3
1.2 Réglementation en matière de rejet des eaux usées.....	4
1.2.1 États Unis.....	4
1.2.2 Aux Pays Bas.....	5
1.2.3 Le Danemark.....	6
1.2.4 Le Royaume-Uni.....	6
1.2.5 L'Allemagne.....	6
1.2.6 Au Québec.....	6
1.3 Étude de sensibilité.....	7
1.4 Les systèmes environnementaux affectés par le débordement.....	8
1.4.1 Cours d'eau récepteur.....	8
1.4.2 La vie aquatique.....	9
1.4.3 Les contraintes.....	9
1.4.4 Charge polluante.....	10
1.4.5 La réaction de l'eau.....	10
1.4.6 La réaction de la vie aquatique.....	11
1.5 Les types et sources de pollution.....	11
1.5.1 Types de polluants.....	11
1.5.2 Les sources.....	14
1.5.3 Facteurs influents sur la qualité des eaux de ruissellement....	15
1.5.3.1 Facteurs reliés à la pluviométrie.....	15
1.5.3.2 Facteurs reliés aux paramètres hydrauliques.....	15
1.6 Phénomène du premier lavage «first flush ».....	16
1.7 Évolution des métaux lourds et contrôle de la source.....	20
1.8 Critères de qualité de l'eau selon le ministère.....	21
1.9 Calcul des objectifs environnementaux de rejet.....	22



CONCLUSION ET RECOMMANDATIONS.....	109
BIBLIOGRAPHIE.....	111

## LISTE DES TABLEAUX

	Page
1.1 Différentes zones de l'effet du premier flot en fonction du paramètre b.....	19
3.1 Efficacité et performance des filtres utilisés.....	53
3.2 Efficacité et performance des filtres utilisés.....	54
3.3a Test de nullité des coefficients d'autocorrélation des résidus pour le pollutogramme.....	55
3.3b Test de nullité des coefficients d'autocorrélation des résidus pour le pollutogramme.....	56
3.3c Test de nullité des coefficients d'autocorrélation des résidus pour l'hydrogramme.....	57
3.3d Test de nullité des coefficients d'autocorrélation des résidus pour l'hydrogramme.....	58
3.4 Coefficient d'autocorrélation des concentrations MES.....	59



4.2a	Reconstitution des données manquantes simulées pour le pollutogramme synthétique (ARIMA).....	105
4.2b	Reconstitution des données manquantes simulées pour le pollutogramme synthétique (interpolation linéaire).....	106
4.3a	Reconstitution des données manquantes. Cas d'un pollutogramme réel.....	107
4.3b	Reconstitution des données manquantes. Cas d'un pollutogramme réel.....	108

## LISTE DES ABRÉVIATIONS ET SIGLES

MES	Matière En Suspension
CUM	Communauté Urbaine de Montréal
USEPA	United States Environmental Protection Agency
NPDES	National Pollutant Discharge Elimination System
CWA	Clean Water Act
MENVIQ	Ministère de l'Environnement du Québec
OER	Objectifs Environnementaux de Rejet
CTAA	Les Critères de Toxicité Aquatique Aiguë
CTAC	Le critère de toxicité aquatique chronique
CCOA	Critère Contamination d'Organismes Aquatiques
$C_E$	La concentration tolérable à l'effluent.
$Q_E$	Le débit de l'effluent
$Q_D$	Le débit alloué pour la dilution
$C_{AM}$	La concentration amont
$Q_{AM}$	Le débit amont
$DBO_5$	Demande biologique en oxygène
HAP	Hydrocarbures aromatiques polycycliques
BPC	Biphényles polychlorés
RL	Filtre à Régression Linéaire
RQ	Filtre à Régression Quadratique
FM	Filtre Médian
WIN	Filtre Winsoriseur
FRAS	Filtre de Fraser
FRÉQ	Filtre Fréquentiel
MISP	Mutually interactive state parameters estimation
$\eta$	Efficacité
P	Coefficient de performance
AR	Modèle Autorégressif
FK	Filtre de Kalman
ARMA	Modèle Autorégressif à moyenne mobile
ARIMA	Modèle Autorégressif non stationnaire à moyenne mobile

## INTRODUCTION

Pour les municipalités, contrôler la pollution des milieux récepteurs qui est due à des débordements dans les réseaux d'assainissement est une tâche ardue qui est non seulement complexe mais aussi onéreuse. La qualité des eaux déversées, la fréquence des débordements, et l'efficacité des méthodes de traitement sont toutes des variables qui ont une influence sur les programmes de régulation. Ces variables ne devraient pas être négligées dans la prise de décision.

En prenant conscience des impacts que peut avoir la pollution des milieux récepteurs sur la santé publique, la vie aquatique, et autres activités (récréatives, économiques), les autorités ont commencé à revoir leurs réglementations en matière de contrôle de la pollution engendrée par les débordements dans les réseaux d'assainissement en temps de pluie. Ces réglementations diffèrent d'un pays à l'autre selon le degré de pollution et le type (lac, cours d'eau) du milieu récepteur. Cependant, malgré les avancées en matière de réglementation, seul le traitement des eaux en temps sec est universellement exigé.

Le cas des débordements en temps de pluie reste peu réglementé pour des raisons d'adéquation des ressources financières et de techniques disponibles.

La communauté urbaine de Montréal est, quant à elle, responsable des eaux usées au moment où elles sont captées dans les intercepteurs, de leur cheminement et des différentes phases de leur épuration. La Communauté Urbaine de Montréal possède un réseau d'intercepteurs qui dessert au total 74 bassins de drainage. Ces derniers sont de deux types, soit les bassins de type séparatif et les bassins de type unitaire. Les bassins de type séparatif sont au nombre de 26 et se trouvent essentiellement dans la partie

ouest de Montréal. La plupart des bassins de type unitaire, au nombre de 48, sont raccordés aux intercepteurs via des ouvrages de régulation qui permettent lors, d'événements pluvieux intenses, de contrôler le débit acheminé vers la station d'épuration, en fonction de la capacité hydraulique des intercepteurs. L'exemple où l'on trouve le plus grand nombre de municipalités est celui du bassin Saint-Pierre qui recueille les eaux usées de dix municipalités situées au centre sud de l'île de Montréal.

## **CHAPITRE 1**

### **CONTEXTE GÉNÉRAL DU PROJET**

#### **1.1 Introduction**

Pour développer un programme, le gestionnaire doit dresser un inventaire des systèmes environnementaux qui seront favorablement influencés par le contrôle des débordements des égouts combinés en temps de pluie. Toutefois, la mise au point d'un programme de contrôle des déversements d'eaux non-traitées directement dans les cours d'eau, causés par des débordements en temps d'orage, en est à ses premiers développements.

L'élimination des rejets causés par les eaux sanitaires, sur le territoire de la CUM, est très avancée en raison du raccordement de la très grande majorité des réseaux d'égout au système d'interception de la station d'épuration.

On entend par surverse tout débordement d'eaux usées dans un cours d'eau sans traitement préalable à la station d'épuration. L'objectif final est le respect des normes en matière de rejet. Pour atteindre cet objectif il est nécessaire de fournir aux gestionnaires

des méthodes d'estimation est filtrage de données de débit et de concentration en matière en suspension.

## 1.2 Réglementation en matière de rejet des eaux usées.

### 1.2.1 États Unis

L'USEPA (United States Environmental Protection Agency) a fait récemment deux pas importants dans le contrôle des eaux d'orage dans les zones urbaines (Roesner et Traina, 1994). Le premier est le développement final de la réglementation et la régulation des réseaux séparatifs. Publiée au début des années 90, la loi exige que toute municipalité dont la population dépasse 100,000 habitants postule pour un permis de déversement des eaux pluviales selon le protocole NPDES (National Pollutant Discharge Elimination System). Le processus d'obtention du permis doit inclure, entre autres, un plan de réduction des charges polluantes jusqu'à «la quantité maximale faisable ».

Le second pas est la publication, en janvier 1993, de l'ébauche d'une politique de régulation des débordements des réseaux combinés qui réaffirme la stratégie de contrôle adoptée en 1989:

- a. Aucun débordement n'est toléré en temps sec;
- b. Utiliser au mieux la technologie et les moyens disponibles pour épurer les eaux combinées selon les exigences du CWA (Clean Water Act);
- c. Minimiser l'impact des débordements des réseaux unitaires sur la santé de l'homme et l'environnement.

La réglementation de 1993 va plus loin en spécifiant les lignes directrices suivantes:

- a. Caractériser précisément la nature des effluents des eaux combinées;
- b. Implanter au minimum neuf contrôles dans le réseau (Roesner et Traina, 1994);

- c. Développer une politique de contrôle à long terme des débordements des réseaux combinés.

Les neuf points minimaux de contrôle sont:

- a. Programme régulier d'entretien, de maintenance et d'opération du réseau et des points de débordement;
- b. Usage maximal du système de collecte pour le stockage;
- c. Révision et modification des programmes de pré-traitement pour minimiser l'impact de débordements;
- d. Maximisation des volumes acheminés à la station d'épuration;
- e. Aucun débordement en temps sec;
- f. Contrôles des matériaux solides et flottants dans les rejets;
- g. Programmes de réduction de la pollution à la source;
- h. Information du public sur les fréquences de débordements;
- i. Caractérisation des effluents pour le suivi des impacts des déversements et l'efficacité des interventions.

Le plan de contrôle à long terme offre aux municipalités le choix entre une approche de présomption fixant la fréquence admissible de débordements (4 à 6 / année) et une approche plus flexible de démonstration consistant à prouver qu'un mode de contrôle donné répond aux critères du CWA (Clean Water Act);

### 1.2.2 Aux Pays-Bas

Les fréquences des déversements autorisés sont comprises entre 3 et 10 par année selon la décision des autorités.

### 1.2.3 Le Danemark

Les recommandations danoises relatives à la qualité des eaux de rivières affectées par des rejets urbains en temps de pluie peuvent être traduites en terme d'acceptation de mortalité piscicole pour une période de retour donnée variant de 8 à 16 ans.

### 1.2.4 Le Royaume-Uni

Les standards de rejet sont définis avec une fréquence de 95% ,ce qui veut dire pas de rejet pour 95% des averses. Ils sont validés et ajustés grâce à des modèles d'estimation des charges polluantes des rejets et de simulation de leur impact sur le milieu récepteur. Ces standards déterminent des teneurs en oxygène dissous en fonction de la période de retour et de la durée des rejets.

### 1.2.5 L'Allemagne

Les prescriptions allemandes s'intéressent plutôt aux exigences en matière de rejets. Certaines normes fixent les rejets d'eaux usées traitées en station. Les Allemands préconisent, de surcroît, un stockage de 25m<sup>3</sup> par ha imperméabilisé, soit 2,5mm de hauteur d'eau.

### 1.2.6 Au Québec

Le Ministère de l'environnement et du Québec admet un débordement par mois dans les zones d'écoulement (cours d'eau), et par deux mois dans les zones d'accumulation. Les débordements ne sont pas tolérés à moins de un kilomètre en amont d'une prise d'eau potable ou d'une zone de cueillette de mollusques. Le règlement 87 de la Communauté Urbaine de Montréal (CUM) prévoit des normes qualitatives et quantitatives pour les principaux polluants de l'eau. Ce règlement exige, entre autres,



que toute personne qui déverse plus de 4500 mètres cubes d'eaux usées industrielles dans un réseau d'égouts durant une période de six mois ou qui déverse des eaux usées contenant des métaux lourds, des cyanures ou des huiles minérales, obtienne un permis de déversement.

Ainsi les structures de régulation des ouvrages d'interception, lesquelles ont été construites sur les conduites d'égout combiné, sont considérées comme des ouvrages de surverses parce qu'elles sont susceptibles de déverser les eaux usées aux cours d'eau. Seize à dix-huit débordements se produisent en moyenne sur l'intercepteur Nord de la CUM du premier mai au 31 octobre et en temps de pluie.

Une certaine quantité d'eau non traitée est rejetée dans le cours d'eau durant des périodes qui varient de quelques minutes à plus d'une journée dépendant de la rapidité avec laquelle se produit la fonte des neiges et le ruissellement pluvial.

### 1.3 Étude de sensibilité

L'étude des sensibilités autour des rives du territoire de la CUM a été réalisée en deux étapes :

- a. Recueillir, analyser et faire une synthèse des connaissances actuelles disponibles sur les ressources biophysiques et les usages de l'eau;
- b. Évaluer la sensibilité des ressources biophysiques et les usages aux rejets d'eaux usées.

Les ministères des affaires municipales et de la métropole et de l'environnement de Québec et la CUM établissent actuellement des exigences et des objectifs de débordements pour la totalité des ouvrages des surverses.

Pour chaque ouvrage des objectifs environnementaux de rejets ont été établis se basant sur une grille d'évaluation élaborée par le ministère de l'environnement du Québec dans la perspective de minimiser les contraintes de récupération des usages de l'eau à partir d'un historique des trois dernières années. Les objectifs spécifiques de rejet vis-à-vis des débordements d'orage sont définis à deux niveaux:

- a. La protection et l'amélioration de l'équilibre de l'écosystème;
- b. La protection et de la récupération des usages.

À ces objectifs spécifiques s'ajoutent des exigences de base qui sont applicables à toutes les surverses. Ces derniers se résument comme suit :

- a. Aucun débordement en temps sec (sauf lors de fontes subites au printemps et de situations d'urgence) ;
- b. Aucune augmentation des débordement en temps de pluie par rapport à la situation actuelle;
- c. Aucun débordement d'eaux usées ayant des fortes charges en DBO<sub>5</sub>, toxiques ou à couleur prononcée.

Les études diverses et de sensibilité fournissent les éléments clé à l'optimisation hydraulique (ouvrages de retenue et priorisation des points de rejets) ce qui rend indispensable la réalisation d'ouvrages de rétention, afin d'atteindre les objectifs cités plus hauts.

#### 1.4 Les systèmes environnementaux affectés par les débordements

##### 1.4.1 Cours d'eau récepteur

Les paramètres qui caractérisent la vulnérabilité de la ressource eau par rapport aux débordements sont:

- a. Débit du milieu récepteur par rapport à la charge déversée;
- b. Usages bénéfiques (histoire, présente, et futur);
- c. Changements saisonniers (écoulement, température, couverture de glace);
- d. Facteurs physiques (pente, vitesse, zones de mélange, regroupement...).

#### 1.4.2 La vie aquatique

La caractérisation de la vie aquatique aide à déterminer les polluants concernés. Les poissons les plus sensibles doivent être identifiés pour développer une approche en vue de leur protection. Les aspects pris en compte sont:

- a. Aspects saisonniers;
- b. Variation passagère de l'oxygène dissous et tout autre produit toxique;
- c. Type et diversité d'autres organismes nécessaires;
- d. Importance relative de la physiologie du cours d'eau.

Les organismes en bas de la chaîne alimentaire sont plus importants à considérer que les poissons et autres vies aquatiques.

Les précipités solides contiennent des quantités importantes de métaux lourds et autres toxiques. De plus le dépôt de particules inertes peut couvrir les organismes en bas et affecter l'environnement.

#### 1.4.3 Les contraintes

L'une des contraintes est l'effet des dépôts inférieurs dus au débordement des égouts combinés qui se sont accumulés précédemment et qui sont classifiés comme toxiques ce qui complique leur enlèvement et leur disposition en lieu sûr. Les débordements municipaux et industriels, les eaux usées, les débits d'agriculture, reste une source de toxiques dans l'alimentation de la chaîne alimentaire.

La physiologie du cours d'eau peut être une contrainte extrêmement importante pour maintenir certaines poissonneries. Les facteurs sont : la vitesse, la température de l'eau, les structures inférieures, les barrières à la migration.

Les usages récréatifs peuvent être contraints par le trafic, les accès limités, la physiologie du cours d'eau, la charge polluante.

#### 1.4.4 Charge polluante

La charge polluante est estimée habituellement sur une base horaire, pour les enregistrements historiques des précipitations à long terme. Cette caractérisation permet aussi la considération de charges annuelles des polluants et la sélection des événements chargés. La caractérisation continue de la charge va mener à une évaluation plus précise des impacts des débordements des égouts combinés en temps de pluie et définir la voie à suivre pour rencontrer les objectifs de diminution de la pollution.

#### 1.4.5 La réaction de l'eau

L'impact des débordements des égouts combinés sur l'eau peut être classifié comme transitoire ou à long terme. Les impacts transitoires ou à effet de choc sont importants quand les standards de qualité sont atteints ou quand les contraintes à court terme sur les poissons et autres vies aquatiques sont importantes. Un événement de «design» ou une caractérisation continue de la charge polluante peut être utilisée pour prédire les impacts transitoires. Les impacts à long terme sont les préoccupations majeures car les polluants sont souvent associés avec les dépôts solides. Les données requises pour prédire la réponse de l'eau peuvent être plus ou moins étendues selon les objectifs fixés. Elles sont sensés comprendre les courants d'eau et le mouvement de sédiments.

#### 1.4.6 La réaction de la vie aquatique

Lors de la caractérisation de l'eau pour déterminer la réaction de la vie aquatique, les besoins des données à propos des formes de vie sont nécessaires. Ces informations doivent refléter comment les formes de vie répondent à des niveaux variables de contraintes et quels sont les paramètres critiques à prendre en compte dans la prédiction des niveaux de contraintes. En temps de pluie, les débordements d'eaux usées peuvent entraîner de brusques modifications du milieu aquatique. Celles-ci peuvent être de nature :

- a. Biologique (augmentation de la charge organique, enrichissement en substances nutritives, contamination bactérienne);
- b. Physique (augmentation de la charge des matières en suspension);
- c. Toxique (métaux lourds);
- d. Esthétique (turbidité du panache de diffusion et débris lavagets);
- e. Olfactive (mauvaises odeurs).

### 1.5 Les types et sources de pollution

#### 1.5.1 Types de polluants

Les paramètres de la pollution peuvent être classés en cinq catégories différentes selon leur action sur la qualité de l'eau ( Ribstien et Desbordes, 1978 ) : Solides en suspension, matière organique, nutriments, métaux lourds et Bactéries.

La mesure de la qualité doit avoir pour but de définir l'impact d'un rejet sur l'environnement. Chaque paramètre doit être étudié en fonction de ses effets sur la faune et la flore aquatiques et sur les principaux usages de l'eau. Mais l'étude de la qualité ne présente pas forcément la mesure de tous ces paramètres. En fait, pour chaque étude, le nombre de paramètres est défini à partir d'une considération budgétaire, beaucoup plus

que par une étude précise des objectifs à atteindre ou d'un niveau minimum de qualité des eaux à maintenir.

Du point de vue de la composition, plusieurs études ont montré que les polluants des eaux sont essentiellement inorganiques. Ce résultat s'explique par le fait que l'érosion des surfaces perméables non construites, des zones en construction et des zones à développement récent, contribue aux apports très importants de graviers, sables, limons et argiles que l'on trouve dans les eaux de ruissellement urbain. En réalité, ces éléments sont inertes et ne présentent pas une source de pollution pour l'environnement mais le problème c'est qu'ils servent comme support pour la matière organique qui altère la qualité de l'eau d'une manière importante ( Ribstien et Desbordes, 1978 ).

On a vu précédemment que les paramètres de la pollution agissent en même temps par leur concentration et par leur charge. En effet, la variation au cours de la pluie de ces deux paramètres est très importante et elle présente essentiellement trois phases (Kothandaraman, 1972 ) :

- a. Une phase initiale au cours de laquelle on a les conditions de temps sec;
- b. Une seconde phase caractérisée par une augmentation du débit et des concentrations d'un certain nombre de paramètres, soit l'«effet de première chasse»;
- c. Une troisième phase où les concentrations sont diluées par rapport aux conditions initiales en raison du débit élevé.

Cet effet de première chasse s'explique par un lavage des dépôts en surface et une remise en suspension des dépôts dans les conduites. Certaines études ont expliqué le phénomène par un effet du premier lavage, c'est à dire des concentrations maxima de polluants dans les premières minutes de la pluie. Schématiquement, le phénomène se présente par un déphasage entre le pollutogramme et l'hydrogramme et on aura par suite au début de ruissellement une pointe de concentration précédant la pointe de débit.

La sédimentation peut avoir des impacts biologiques importants sur l'habitat de la faune et la flore aquatiques:

- a. Altération d'habitats: l'érosion des berges et les dépôts sur le lit du cours d'eau peuvent empêcher certains organismes de procréer;
- b. Réduction de la capacité des ouvrages : les dépôts de sédiments peuvent réduire l'emménagement des ouvrages de rétention et réduire la section d'écoulement des conduites, augmentant ainsi le risque d'inondation;
- c. Demande en oxygène : la turbidité peut empêcher certains organismes aquatiques d'obtenir de l'oxygène en quantité suffisante pour leur survie;
- d. La diminution de la pénétration de la lumière affecte l'activité photosynthétique;
- e. La présence de sédiments peut augmenter les coûts de traitement pour obtenir de l'eau potable.

Une quantité importante de phosphore et d'azote peut accélérer la prolifération d'algues, ce qui provoque la décoloration de l'eau, le développement d'odeurs fortes et la diminution de l'oxygène dissous.

Les bactéries et virus peuvent causer des maladies surtout lorsqu'il s'agit d'un milieu récepteur à usage récréatif.

La décomposition des matières organiques par les micro-organismes diminue les niveaux d'oxygène dissous dans les eaux à faible courant (lacs, estuaires). Ceci implique une baisse de DBO.

Les métaux lourds et produits chimiques peuvent intoxiquer la vie aquatique et contaminer les sources d'approvisionnement en eau potable.

Les huiles et les graisses ainsi que les hydrocarbonates peuvent être fatals pour la faune aquatique, car ils affectent l'apparence de la surface de l'eau, et ont une tendance à s'agripper aux sédiments au fond des cours d'eau.

Les sels proviennent du déglacage des rues et sont toxiques pour plusieurs organismes aquatiques qui ne peuvent tolérer qu'un certain niveau de salinité.

### 1.5.2 Les sources

La composition chimique des eaux de ruissellement diffère selon le type d'occupation du sol, on distingue quatre types d'occupation : Secteurs résidentiels et commerciaux, secteurs industriels, routes et autoroutes, sites de construction.

D'après le programme américain NURP (Nationwide Urban Runoff Program) les concentrations en polluants les plus importantes sont les MES, DBO, DCO, et métaux lourds. Pour les bassins à occupation commerciale et résidentielle les facteurs les plus importants pouvant influencer les charges polluantes sont le pourcentage d'imperméabilisation et la quantité de précipitation.

Les quantités les plus grandes et les polluants les plus variés se trouveront dans l'eau de ruissellement provenant des zones industrielles qui se caractérisent par un fort pourcentage de surfaces imperméables. Les ruissellements issus de ces surfaces contiennent des éléments toxiques, parfois à des niveaux importants.

Les concentrations des polluants que l'on trouve dans les eaux de ruissellement des routes principales et autoroutes sont plus élevées que celles provenant des zones résidentielles et commerciales. Il est aussi établi que les charges polluantes des routes en zone urbaine sont plus importantes que celles des zones rurales car le nombre de véhicules est plus important.



Les quantités de sédiments générées sur un site de construction sont vingt fois plus élevées que sur un territoire en culture et mille à deux mille fois plus élevées que par une forêt.

### 1.5.3 Facteurs influant sur la qualité des eaux de ruissellement pluvial

L'étude du ruissellement urbain exige, outre la connaissance des sources de la pollution, une limitation des facteurs influant sur la qualité des eaux pluviales. En fait ces facteurs présentent deux classes importantes ( Philippe, 1984 ) :

#### 1.5.3.1 Facteurs reliés à la pluviométrie.

Ces facteurs sont :

- a. La hauteur locale d'averse;
- b. L'intensité moyenne maximale, sur une durée caractéristique contribuant à la formation de l'hyétogramme. Le temps de concentration est généralement adopté comme durée caractéristique;
- c. La durée de l'averse;
- d. La durée du temps sec précédant l'averse.

#### 1.5.3.2 Facteurs reliés aux paramètres hydrauliques

Ces facteurs sont :

- a. Le volume du ruissellement pluvial ou le volume total qui tient compte du volume du ruissellement de temps sec;
- b. Le débit maximal de pointe des eaux pluviales et le débit maximal instantané d'effluent d'orage qui est la somme de deux débits en temps sec et en temps humide.

- c. Les événements les plus importants en charge et en concentration moyenne, ne peuvent s'expliquer simplement que par une caractéristique déterminée indépendamment des actions des autres. Ces événements polluants se produisent lors des situations bien particulières, résultant de la combinaison de plusieurs caractéristiques hydrauliques et climatiques.

#### 1.6 Phénomène du premier lavage «first flush»

La partie la plus polluante des eaux de ruissellement est celle du début de la pluie qui lessive les rues, les toits et les conduites d'égout pluvial. Elle est appelée communément le «first flush» et est définie jusqu'à présent de façon subjective. En effet, le lessivage réalisé par cette première pluie est lié à plusieurs facteurs présentant une grande variabilité tels que la superficie du bassin, son imperméabilité, et l'occupation de son sol. On a relevé certaines définitions (Saget, 1996) que l'on peut considérer représentatives du «first flush» (Thornton et Saul, 1987; Geiger, 1987; EPA (1993); Krajewski et al, 1995 et 1998). Celles qui nous semblent pertinentes sont de:

- a. Thornton et Saul (1987), qui définissent le premier lavage comme étant le volume de ruissellement précède la pointe le pic de la concentration des matières en suspension;
- b. L'EPA (1993), qui définit le premier lavage comme étant le volume de ruissellement écoulé entre les instants  $t_1$  et  $t_2$  où la concentration devient respectivement supérieure et inférieure à la concentration en temps sec.

Geiger (1987) , pour connaître l'existence ou l'absence du phénomène du premier lavage représente graphiquement la distribution de la charge polluante en fonction du volume de ruissellement :

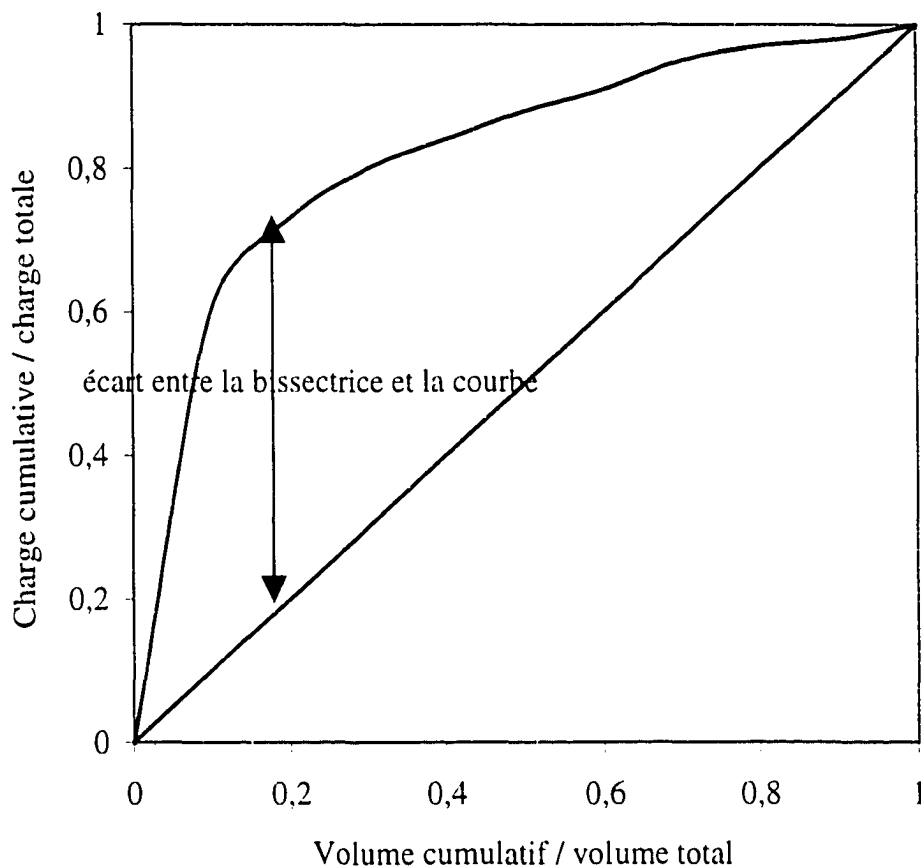


Figure 1.1 Charge cumulative en fonction du volume cumulatif.

- Si la charge polluante est constante pendant l'événement la courbe sera confondue avec la diagonale ;
- Si la charge polluante est supérieure au début de la pluie alors la courbe sera au dessus de la diagonale;
- Si la charge polluante est supérieure à la fin de la pluie alors la courbe sera au-dessous de la diagonale.

Geiger (1987) avance que le phénomène du premier lavage existe si l'écart entre la bissectrice et la courbe est de 0.2. les équations qui représentent mathématiquement la courbe sont :

$$y(j\Delta t) = \frac{\sum_{i=1}^j M(i\Delta t)}{M_{\text{totale}}} = \frac{\sum_{i=1}^j C_i \times Q_i \times \Delta t}{\sum_{i=1}^N C_i \times Q_i \times \Delta t} = f \left( \frac{\sum_{i=1}^j Q_i \times \Delta t}{\sum_{i=1}^N Q_i \times \Delta t} \right) = f \left( \frac{\sum_{i=1}^j V_i(i\Delta t)}{V_{\text{totale}}} \right) = f(x(j\Delta t)) \quad (1.1)$$

où

$y(j\Delta t)$  : représente à l'instant  $j\Delta t$ , la fraction de la masse totale d'un polluant;

$x(j\Delta t)$  : représente à l'instant  $j\Delta t$ , la fraction du volume total véhiculé;

$M(i\Delta t)$  : est la masse d'un polluant donné véhiculée pendant l'intervalle de temps  $i\Delta t (i=1, \dots, j)$ ;

$C_i$  : la concentration du polluant pendant l'intervalle  $i\Delta t$ ;

$N$  : est le nombre de points d'échantillonnage dans le temps;

$\Delta t$  : est le pas de temps d'échantillonnage.

Gupta et Saul (1996) proposent de tracer les courbes de volumes et de charges cumulatives en fonction du temps et de déterminer le phénomène du premier lavage comme étant l'écart maximal entre les deux courbes (Figure 1.1).

Krajewski et al (1995, 1998) proposent de remplacer (l'équation 1.1) par la relation suivante :

$$y(j\Delta t) = \frac{\sum_{i=1}^j M(i\Delta t)}{M_{\text{totale}}} = \left( \frac{\sum_{i=1}^j V_i(i\Delta t)}{V_{\text{totale}}} \right)^b = (x(j\Delta t))^b \quad (1.2)$$

La valeur du paramètre  $b$  indique la présence ou l'absence du phénomène du premier flush (Tableau 1.1). La détermination des valeurs de  $b$  se fait par simple régression linéaire à partir de :

$$LN(y(j\Delta t)) = LN\left(\frac{\sum_{i=1}^j M(i\Delta t)}{M_{totale}}\right) = b \times LN\left(\frac{\sum_{i=1}^j V_i(i\Delta t)}{V_{totale}}\right) = b \times LN(x(j\Delta t)) \quad (1.3)$$

Tableau 1.1

Différentes zones de l'effet du premier lavage en fonction du paramètre b.

Valeur de b	Zone	Importance de l'effet du premier lavage	
$0 < b < 0.185$	1	Positif	prononcé
$0.185 < b < 0.862$	2	//	moyen
$0.862 < b < 1.000$	3	//	faible
$1.000 < b < 1.159$	4	Négatif	faible
$1.159 < b < 5.395$	5	//	moyen
$5.395 < b < +\infty$	6	//	Prononcé

Le paramètre b étant directement relié à la distance entre la courbe M(V) et la bissectrice, cette formulation est similaire à la précédente. (L'équation 1.2) possède l'avantage de mieux préciser l'absence ou la présence du premier lavage à travers la valeur d'un paramètre unique et précis : b.

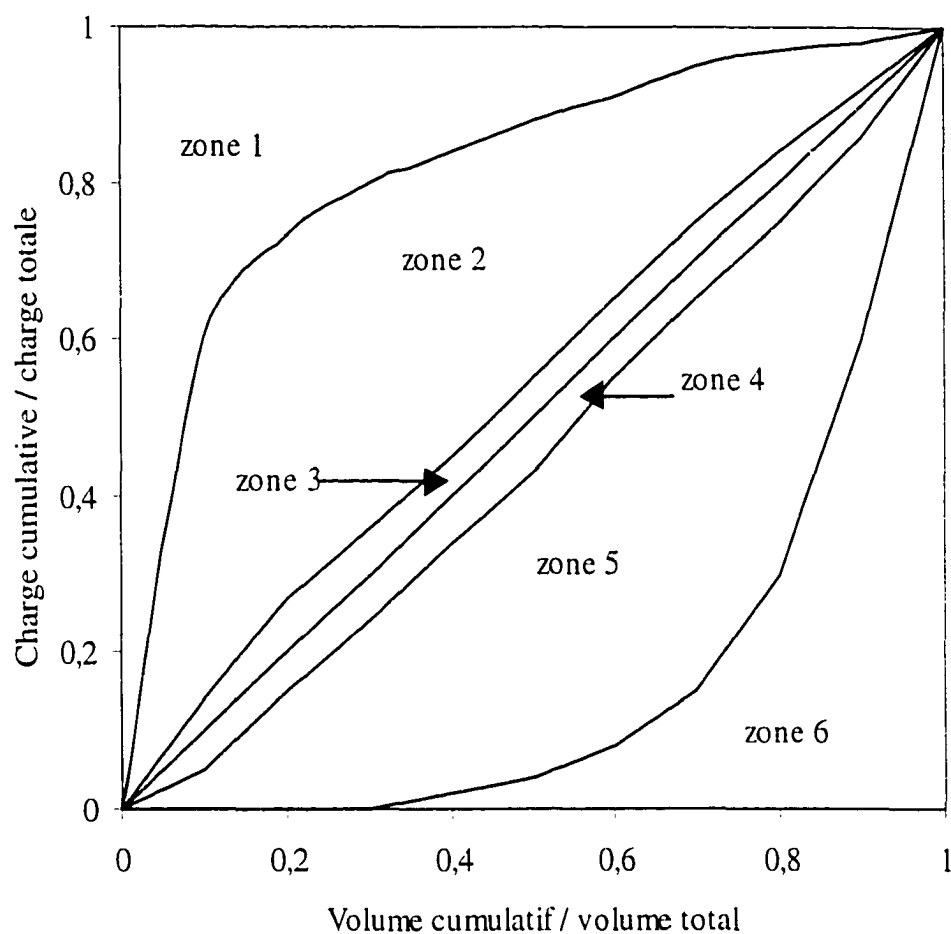


Figure 1. 2 Les différentes zones des courbes  $M(V)$  en fonction du paramètre  $b$

### 1.7 Évolution des métaux lourds et contrôle de la source à la CUM

On dénombre sur le territoire de la CUM plus de 8500 établissements industriels, commerciaux et institutionnels susceptibles de générer des eaux usées. Toutefois, 1200 établissements déversent chacun, annuellement, 9000 mètres cubes ou plus d'eaux usées dans les égouts du territoire. Le volume total d'eaux usées industrielles est évalué 187 millions de mètres cubes par années et représente environ 20% du volume total d'eaux usées en temps sec et plus de 25% de la charge des résidus solides traités à la station

d'épuration. Depuis 1977, des programmes relatifs à la dépollution des rejets d'eaux usées dans les réseaux d'égouts ont été demandés aux industries pour s'assurer que leurs déversements d'eaux usées seraient conformes aux exigences du ministère de l'Environnement du Québec et par la suite aux normes du règlement 87 adopté par la communauté urbaine de Montréal en juin 1986. Les polluants industriels sont classés en deux groupes : les polluants compatibles et les polluants incompatibles aux ouvrages d'épuration.

Dans le premier groupe, on retrouve les constituants typiques des eaux usées domestiques comme les matières en suspension et la charge organique.

Les polluants du deuxième groupe sont ceux susceptibles de causer des dommages aux égouts, de nuire au fonctionnement de la station d'épuration ou d'être préjudiciables au cours d'eau récepteur.

Les plus importants sont les métaux lourds (chrome, cuivre, plomb, cadmium, arsenic, nickel, zinc, mercure), les cyanures, les composés phénoliques, les huiles et les graisses, les produits acides ou alcalins, les polluants chimiques organiques tels les hydrocarbures aromatiques polycycliques (HAP) et le biphényles polychlorés (BPC).

#### 1.8 Critères de qualité de l'eau selon le ministère de l'environnement du Québec

Les critères de qualité de l'eau (MENVIQ, rév. 1992) servent de base de calcul des objectifs environnementaux de rejet (OER). Ils peuvent aussi servir comme une référence pour évaluer l'état de santé du milieu récepteur et l'atteinte des objectifs prioritaires fixés par le Ministère de l'environnement du Québec.

Ces critères sont définis pour les contaminants susceptibles de se retrouver dans les effluents municipaux et industriels. Ces critères assurent les objectifs suivants :

- a. La protection de la santé humaine ;
- b. La protection de la vie aquatique incluant la faune terrestre associée;
- c. La protection des activités récréatives.

Le critère «vie aquatique » d'un contaminant représente la concentration à ne pas dépasser dans le milieu pour protéger les organismes aquatiques des impacts attribuables à ce contaminant. Ce critère est défini pour deux durées d'exposition. Le critère de toxicité aquatique aiguë (CTAA) indique la concentration maximale à ne pas dépasser pour protéger les organismes pendant une exposition de courte durée. Le critère de toxicité aquatique chronique (CTAC) indique la concentration maximale à ne pas dépasser pour protéger les organismes pendant une exposition de longue durée.

Le critère «santé humaine » d'un contaminant représente la concentration à ne pas dépasser dans le milieu pour protéger les individus des impacts néfastes engendrés par la consommation d'eau et d'organismes aquatiques. Ainsi le critère «eau brute » protège à long terme la santé d'un individu qui consomme de l'eau et des organismes aquatiques venant d'un même plan d'eau. D'autre part, le critère «contamination d'organismes aquatiques » protège à long terme la santé d'un individu qui ne fait que consommer des organismes aquatiques venant d'un plan d'eau.

Le critère «activités récréatives » d'un contaminant vise à prévenir les dangers pour la santé liés au contact primaire (baignade) ou au contact secondaire (navigation, pêche, canotage) avec l'eau. En outre ce critère garantit la protection de la qualité esthétique du milieu aquatique.

### 1.9 Calcul des objectifs environnementaux de rejet

La détermination des OER a pour but le maintien et la récupération de la qualité des milieux aquatiques. Les objectifs de rejet peuvent être quantitatifs ou qualitatifs. Ils comprennent aussi une exigence de non-toxicité aiguë à l'effluent qui est vérifiée par la réalisation de bioessais. Les objectifs qualitatifs sont reliés aux qualités esthétiques des plans d'eau. Les objectifs quantitatifs définissent les concentrations et charges des différents contaminants qui devraient être sécuritaires pour le milieu récepteur.



Les OER quantitatifs sont établis pour assurer le respect des critères de qualité de l'eau à la limite d'une zone de mélange restreinte. Une zone de mélange est tolérée car une zone d'impact local ne devrait pas nuire au reste du cours d'eau.

La méthode de calcul des OER est basée sur un bilan de charge simple appliqué sur une portion du cours d'eau. Ce bilan est établi de façon à ce que la charge déjà existante en amont du rejet, à laquelle est ajoutée la charge de l'effluent, respecte la charge maximale admissible à la limite de la zone de mélange. Le calcul des OER intègre donc la qualité des eaux en amont du point de rejet, le débit du cours d'eau en période d'étiage, le débit de l'effluent et les critères de qualité de l'eau permettant d'assurer la protection ou la récupération des usages.

Pour assurer la protection de la vie aquatique, de la faune terrestre et de la santé humaine, le critère «toxicité aquatique chronique (CTAC) » et le critère «contamination d'organismes aquatiques (CCOA) » doivent toujours être respectés à la limite aval de la zone de mélange. Toutefois, si une prise d'eau potable, une plage, une zone de canotage ou tout autre usage spécifique est présent en aval d'un rejet ponctuel, on s'assure que le critère s'appliquant à cet usage sera respecté au point d'usage.

Le bilan de charge est présenté par l'équation suivante :

Charge maximale, à la limite de la zone de mélange = charge amont charge tolérable à l'effluent

$$(Q_D \times \text{critère}) = (Q_{AM} \times C_{AM}) + (Q_E \times C_E). \quad (1.4)$$

où  $Q_D$  est le débit de dilution;

$Q_{AM}$  est le débit amont;

$Q_E$  est le débit de l'effluent;

$C_{AM}$  est la concentration amont;

$C_E$  est la concentration de l'effluent.

La charge tolérable à l'effluent ( $Q_E \times C_E$ ) est donc égale à :

$$(Q_E \times C_E) = (Q_D \times \text{critère}) - (Q_{AM} \times C_{AM}). \quad (1.5)$$

Et la concentration tolérable à l'effluent ( $C_E$ ) de l'industrie est égale à :

$$(C_E) = \frac{(Q_D \times \text{critère})}{Q_E} - \frac{(Q_{AM} \times C_{AM})}{Q_E}. \quad (1.6)$$

La concentration amont ( $C_{AM}$ ) de chaque paramètre est déterminée à partir de données représentatives de la qualité du cours d'eau en amont du rejet.

Le débit de l'effluent ( $Q_E$ ) correspond généralement au débit annuel moyen de l'effluent.

Le débit alloué pour la dilution ( $Q_D$ ) est la somme du débit de l'effluent ( $Q_E$ ) et du débit amont, ( $Q_{AM}$ ), ce dernier étant la qualité d'eau du milieu allouée pour la dilution. La détermination de cette qualité d'eau est restreinte par les limites physiques imposées à la zone de mélange et celles-ci correspondent à la plus restrictive des définitions suivantes :

- a. Longueur maximale de 300 m, ou moins s'il y a présence d'usages spécifiques autres que la vie aquatique, la faune terrestre ou la consommation de poissons;
- b. Largeur inférieure à la moitié de la largeur du cours d'eau jusqu'à un maximum de 50 m pour maintenir une zone de passage pour les organismes aquatiques;
- c. Dilution maximale de l'effluent de 100 fois dans le débit alloué pour la dilution.

En pratique, dans une petite rivière, le débit maximal alloué pour la dilution est égal à la moitié du débit de la rivière en période critique (jusqu'à l'atteinte de la dilution maximale de 1 dans 100). Alors que dans une grande rivière (plus de 100m

de large), où l'hydrodynamique est plus complexe, l'utilisation d'un test de diffusion est parfois nécessaire pour évaluer la dilution de l'effluent à 300 m du point de rejet.

## **CHAPITRE 2**

### **REVUE DE LITTÉRATURE**

Dans Le chapitre 1 nous avons mis en évidence tous les problèmes que la pollution en milieu urbain peut causer pour le milieu récepteur : la santé publique; la faune et la flore aquatique. Nous allons maintenant nous intéresser à l'étude des MES qui sont les principaux vecteurs de la pollution. Mais le problème qui se pose est que les données recueillies lors des campagnes de mesures des concentrations sont souvent peu fiables (erronées, manquantes). Finalement pour augmenter le degré de confiance accordé à ces mesures nous devons tout d'abord nous assurer de la performance et de l'efficacité des filtres décrits dans la littérature et des procédures qui résultent d'une combinaison gagnante de ces modèles (chapitres 3 et 4). L'objectif principal de ces filtres et de permettre aux gestionnaires des municipalités à travers le Québec, lors de l'acquisition en temps réel des mesures en MES, de valider correctement leurs données et de prendre les meilleures décisions concernant l'acheminement de la partie des eaux fortement polluées vers la station d'épuration, tout en évitant un débordement éventuel. C'est une gestion dite optimale qui permettra de mieux cerner les besoins en matière de réhabilitation des conduites et ouvrages de rétention et les besoins en construction de nouveaux ouvrages.

L'approche de filtrage en temps différé appliquée aux séries de mesures en MES, qui fait l'objet du chapitre 3, a déjà été réalisée dans la littérature pour les données hydriques. En effet, Charbonneau et Bérubé (1987) ont développé une procédure pour la validation des mesures de niveaux amonts journaliers historiques.

Il est question d'un filtre passe bas ayant une fréquence de coupure qui est choisie préalablement selon la nature de la série que l'on veut valider.

que l'on veut valider. Cette méthode élimine les fluctuations aberrantes observées dans les séries. Elle a néanmoins certaines lacunes :

- a. Elle ne tient pas compte d'informations physiques telles que les variables météorologiques ou hydrométriques enregistrées dans les stations avoisinantes;
- b. La fréquence des mesures négatives, bien que sensiblement diminuée par l'application du filtre, reste après validation;
- c. Il n'y a pas de méthodes pour déterminer la fréquence de coupure;
- d. Pour valider la mesure de l'instant «  $t$  », cette technique nécessite la connaissance des mesures futures «  $t+1$  », «  $t+2$  ». On ne peut donc pas appliquer ce filtre en temps réel.
- e. Elle sous-estime les pointes de crues en éliminant les variations de grandes fréquences.

Afin de palier à ces inconvénients Perraut et *al* (1991) ont développé un modèle de validation et d'estimation des données en temps réel. Leur approche est basée sur deux corrections des apports non filtrés :

- a. Une première correction, effectuée a priori, a pour but d'éliminer les mesures improbables qui pourrait biaiser l'estimation des paramètres des modèles utilisés;
- b. La seconde correction consiste à considérer les mesures filtrés par la série de Fourier comme variable dépendante dans une équation de régression.

L'approche Perraut et *al* (1991) est utilisable en temps réel elle permet de corriger l'apport naturel juste après la transmission de la donnée et sans atténuer les pointes de crues. Mais, elle possède un inconvénient qui est l'utilisation des valeurs explicatives douteuses.

Dans le cadre de la validation en temps réel, Roy et *al* (1992) ont suggéré la méthode des niveaux virtuels pour calculer le niveau moyen d'un réservoir à partir de plusieurs

limnimètres situés sur le même réservoir. Ce processus de calcul est itératif et suppose que le niveau virtuel a été calculé pour les jours précédents.

La reconstitution des données manquantes que l'on va traiter dans le chapitre 4 a aussi fait l'objet de plusieurs publications. Bennis et Bruneau (1993 a) ont comparé différentes techniques pour la reconstitution des débits journaliers. En premier lieu la régression multiple qui exprime le débit de la station à reconstituer en fonction des stations voisines (Haan, 1977). Les paramètres de la régression sont déterminés par la méthode standard des moindres carrés (Young, 1974) qui possède trois inconvénients :

- a. Elle ne tient pas compte de la variation des paramètres en fonction du temps;
- b. Les variables indépendantes sont souvent corrélées entre elles, ce qui viole une des conditions de son application;
- c. Les erreurs de mesure obéissent rarement à l'hypothèse de bruit gaussien utilisée dans la solution par la méthode des moindres carrés.

L'algorithme récursif des moindres carrés (Schilling et Martens, 1986) permet de réajuster les coefficients du modèle au fur et à mesure de la réception de nouvelles données. D'autre part pour remédier à la corrélation des variables explicatives, une étude statistique qui porte sur l'analyse en composantes principales a été réalisée (McCuen et Snyder, 1986). En plus de ces techniques qui ont besoin d'information régionale, le modèle autorégressif (Box et Jenkins, 1976), qui tient compte seulement des informations à la station d'intérêt, a été aussi élaboré.

Les techniques utilisées précédemment présentent toutes l'inconvénient de produire des résidus autocorrélés. Pour solutionner ce problème Bennis et Bruneau (1993b) ont utilisé le filtre de Kalman. Le choix de ce filtre est motivé par la précision sur l'estimation des données manquantes et par la production de résidus indépendants. La méthode standard des moindres carrés et deux configurations différentes des moindres carrés ont été comparées. Les résultats montrent effectivement que l'utilisation de ce

filtre produit des résidus indépendants. Pendant cette étude, l'évaluation des bruits du modèle et de mesure reste un point critique.

Bennis, Côté et Kang (1996) ont développé une méthode qui s'applique aux mesures de niveaux en temps différé, à long terme, pour la validation des apports naturels. Le cas étudié est celui où on dispose de plusieurs stations limnimétriques. Les auteurs ont utilisé une méthode appelée multifiltre qui consiste à prévoir la valeur du niveau pour une station donnée à l'aide d'un modèle autorégressif. Entre le niveau mesuré et le niveau estimé par la régression, on choisit comme niveau valide, celui qui est le plus près du niveau prévu par le modèle autorégressif. L'algorithme de validation est effectué dans le sens «Aller» et «Retour» dans le temps. La valeur mesurée est considérée aberrante si elle a été éliminée dans les deux étapes de validation. La technique du filtre de Kalman est utilisée pour estimer les paramètres des deux modèles régressifs et autorégressifs.

Berrada, Bennis, et Gagnon (1996) ont traité la validation des apports naturels dans le cas où les réservoirs seraient équipés d'un seul limnimètre. Des techniques univariées de filtrage ont été utilisées, à savoir : les filtres linéaires, quadratique, médian, winsoriseur à priori et à posteriori (Ragot et al, 1990), fréquentiel (Van Den Enden et Verhoeckx, 1992) et Fraser (Radix, 1984) combiné avec le filtre de Kalman. Ces différents filtres ont été d'abord appliqués sur une série chronologique idéale, laquelle a été bruitée par un bruit blanc et puis par un bruit de grandes amplitudes, ensuite sur des séries historiques réelles de niveaux d'eaux des réservoirs. À l'aide des critères qualitatifs et quantitatifs, les auteurs ont comparé la performance des techniques utilisées. Les techniques de filtrage utilisées ont donné dans l'ensemble des résultats satisfaisants. Le filtre de Fraser a été suggéré pour détecter les valeurs aberrantes isolées, tandis que les filtres fréquentiel, médian, élabreur winsoriseur, linéaire et quadratique, sont appropriés pour valider les séries très bruitées.

Bennis, Berrada et Kang (1997) ont développé une technique fortement efficace pour obtenir la meilleure approximation des moindres carrés des données hydrologiques manquantes dans le cas univarié. La technique est basée sur le poids approprié des valeurs estimées produites par deux processus auto-régressifs fonctionnant, respectivement, dans les directions aller et retour dans le temps. Pour le cas multivarié deux modèles de régression multivariés sont calibrés périodiquement sur des données disponibles avant et après la période des données manquantes. L'utilisation du filtre de Kalman a amélioré l'exactitude dans l'évaluation des premières données manquantes comprenant le débit de pointe. Pour des données manquantes ultérieures la confiance des évaluations est plus grande en utilisant un modèle statique identifié par la technique ordinaire des moindres carrés. On a constaté qu'il y a un grade critique pour lequel il y a une inversion d'exécution entre la technique du filtre de Kalman et la méthode ordinaire des moindres carrés.

Fich et McBean (1991) ont développé un modèle pour la prévision des écoulements de fleuve résultant de la fonte de neige et de la précipitation. Le modèle utilise un filtre de Kalman pour refléter l'incertitude dans les données mesurées et des paramètres du modèle de système. Les auteurs ont conclu que l'algorithme est montré pour développer de bonnes prévisions d'une journée et de deux jours, mais le modèle linéaire de prévision est insatisfaisant pour des prévisions à plus long terme. Ils ont recommandé que de bonnes évaluations initiales de paramètres s'avèrent essentielles pour l'exécution optimale des prévisions.

Enfin, Bennis, Berrada et Bernard (article sous presse) ont élaboré une méthodologie de validation des données hydrométriques mesurées dans un réseau d'assainissement urbain. Ils ont adopté une méthode multivariée, basée sur la redondance analytique. La démarche suivie est identique à celle de Bennis, Côté et Kang (1996) sauf que celle-ci est basée sur la redondance matérielle en temps différé. La méthodologie proposée a



permis de corriger les aberrations de la série des mesures préalablement bruitée par un bruit blanc et puis par un bruit de grande amplitude.

Lors de cette étude, dans une première étape on applique dans un environnement univarié différents filtres et combinaisons de filtres à des séries de mesures synthétiques bruitées afin de trouver la combinaison la plus performante en terme de filtrage de données. Ensuite on l'applique sur des données réelles en vue d'obtenir une représentation optimale des séries en question.

## **CHAPITRE 3**

### **FILTRAGE UNIVARIÉ DES MESURES DANS UN SYSTÈME D'ASSAINISSEMENT**

#### **3.1 Introduction**

Compte tenu de la vétusté des réseaux d'assainissement, de l'urbanisation, et de l'augmentation corrélative des surfaces imperméables, de plus en plus de municipalités sont aux prises, en temps de pluie, avec de graves problèmes d'inondation. Le milieu naturel n'est guère épargné; les débordements fréquents d'eaux pluviales mélangées avec des eaux usées dans les réseaux unitaires ont fait perdre aux rivières, lacs et fleuves la plupart de leurs usages tout en menaçant la santé publique.

L'essor des technologies de l'information est à la base d'un renouveau dans la gestion dynamique en mode de contrôle prédictif intégré des réseaux d'assainissement. La centralisation des données recueillies en temps réel sur la qualité et la quantité des flux à différents endroits du réseau permet d'agir à distance. Malheureusement, les données recueillies sont toujours bruitées et parfois fausses ou manquantes.

Les sources d'erreurs dans la mesure des écoulements sont très nombreuses. À titre d'exemples, on peut citer les erreurs qui proviennent de l'encrassage et du colmatage de

la tête d'un capteur ou d'une crépine, d'une dérive ou d'un mauvais étalonnage ou encore d'un problème de transmission qui génère une série de données manquantes (Chocat, 1997). Il peut s'agir d'un curage, d'un arrosage, d'une déviation à l'amont, d'un rejet accidentel ou d'une accumulation de dépôt qui crée un barrage (Piatyszek, 1998). En se basant sur les travaux de Winkler et Makridakis (1983), Perreault et al. (1995) proposent une méthode de validation basée sur la combinaison de deux composantes, spatiale et temporelle. Les coefficients de pondération sont calculés de manière à minimiser la somme des carrés des écarts entre les valeurs validées et les valeurs obtenues par un filtre « fréquentiel » passe-bas. Cette méthode possède les lacunes suivantes :

1- Toutes les valeurs mesurées sont systématiquement modifiées alors que notre but est de corriger uniquement les valeurs réellement aberrantes : 2- Les valeurs de référence utilisées dans le processus de validation sont celles filtrées par une méthode univariée utilisant une fréquence de coupure choisie arbitrairement. Pour valider les niveaux mesurés en temps réel, Roy et al. (1992) ont proposé la méthode des niveaux virtuels. Cette approche qui fournit d'assez bons résultats possède les mêmes inconvénients que l'approche précédente. Nguyen et Bisson (1998) ont comparé le modèle de lissage exponentiel, au modèle de Perreault et al. (1995), et ont conclu que le premier est plus performant selon les critères d'écart quadratique moyen et de conservation des volumes. Le modèle de lissage qui a l'avantage d'être simple de mise en œuvre possède les inconvénients des modèles précités. Quatre types de problèmes, que nous rencontrons couramment dans les données ont été traités au sein de notre équipe : 1. La validation dans un environnement univarié (Berrada et al, 1996) pour lequel, une seule station de mesure existe (système simplex). 2. La validation dans un environnement multivarié (Bennis et Kang, 2000) pour lequel il y a redondance matérielle (système multiplex utilisant plusieurs appareils de mesure placés au même endroit pour mesurer la même grandeur). 3. L'estimation des données manquantes dans un environnement univarié (Bennis et al., 1997). 4. L'estimation des données manquantes dans un environnement multivarié (Bennis et al., 1997).

Cette recherche qui constitue indéniablement une percée significative dans le domaine de la validation se limite, cependant, aux données hydrométriques journalières en milieu rural. La validation des données hydrométriques en milieu urbain a été peu documentée dans la littérature, tandis qu'une méthodologie de validation des mesures à des concentrations de polluants est pour toute fin pratique inexistante. Ceci s'explique par des appareils de mesure, in situ, de concentrations de polluants qui ne sont apparues sur le marché que récemment. Piatyszek (1998) propose une technique de validation des mesures en réseau d'assainissement basée sur le test de Wald (1947). Il s'agit en fait d'une analyse statistique de la variation du débit destinée à renseigner si les conditions d'écoulement en réseau correspondent à un temps sec ou pluvieux. D'autres tentatives de validation des données se sont arrêtées au niveau de contrôle des dépassements d'échelle, des défauts de capteur et des dérives ou pannes de courant (Krajewski, 1996).

Très récemment, Bennis et al. (2001) ont proposé une méthodologie de validation des débits mesurés dans un réseau d'assainissement en utilisant la redondance analytique. Cette méthodologie de validation consiste à comparer le débit mesuré au débit estimé à l'aide d'un modèle hydrologique bien calibré. Un processus de vote logique utilisant un modèle autorégressif comme indicateur de la tendance de variation permet de choisir la valeur la plus vraisemblable parmi les valeurs mesurées et estimées.

Le cas traité dans ce travail correspond à la situation où il y a absence de redondances matérielle et analytique. Dans un souci de clarté, le travail a été décomposé en trois parties.

La première partie, théorique, présente un large éventail de techniques de filtrage qui s'étend du filtre simple utilisant la médiane ou la moyenne mobile, jusqu'au lisseur de Fraser couplé à celui de Kalman. Dans la deuxième partie on applique les techniques de filtrage sur des séries synthétiques. Deux séries chronologiques simulées, une relative au débit et l'autre à la concentration en matières en suspension, ont été artificiellement

bruitées à la fois par un bruit blanc et par un certain nombre de perturbations de grandes amplitudes. L'application des filtres sur ces séries permettra de tirer des conclusions sur l'opportunité d'application des différentes méthodes. La troisième partie permettra d'appliquer les filtres utilisés sur des séries réelles mesurées. Les critères d'efficacité et de performance introduits sont largement concluants pour les données réelles et synthétiques.

### 3.2 Description des techniques univariées de filtrages

#### 3.2.1 Introduction

Un bassin hydrologique, rural ou urbain, joue un rôle naturel de filtre dont l'entrée est la pluie et la sortie est l'hydrogramme et le pollutogramme des différents polluants. Lorsqu'il n'y a pas d'intervention extérieure, telle qu'une manipulation de vanne ou une décharge non habituelle de polluants dans le réseau d'assainissement, la variation temporelle d'une donnée manifeste normalement une certaine persistance au cours de l'événement. Ce sont ces tendances haussières ou baissières qui influencent le gestionnaire dans ses prises de décision pour la rétention ou le délestage des flux. Les petites fluctuations courantes et les grandes variations isolées dans les concentrations et les débits peuvent conduire à des incohérences dans la prise de décision. On peut citer à titre d'exemple la consigne d'ouvrir une vanne qu'on a commencé de fermer durant le pas de temps précédent (généralement de cinq minutes) compte tenu d'une variation soudaine dans la concentration d'un polluant.

Plutôt que de faire un lissage intuitif sur les prises de décision, il est recommandé d'appliquer les techniques de filtrage sur les données de base qui alimentent les modèles d'optimisation. Ces données sont les hydrogrammes et les pollutogrammes mesurés à différents endroits d'un système d'assainissement.

En dehors des filtres fréquentiels et de Fraser, tous les filtres classiques que l'on décrit ci-dessous (Berrada et al, 1996), utilisent un échantillon représenté par une fenêtre mobile, comprenant un certain nombre de mesures, fixé a priori. À chaque pas de temps, c'est la donnée centrale d'une fenêtre qui fait l'objet de traitement.

### 3.2.2 Les filtres moyenneurs

Ces filtres qui suivent la tendance centrale (moyenne) d'une série chronologique sont bien adaptés au lissage des bruits de mesure de faible amplitude relative, mais sensibles aux valeurs de grandes amplitudes qui contaminent le signal (Swartz, 1989). Ces filtres s'expriment par la relation suivante :

$$y_t = \sum_{k=-m}^m \alpha_k x_{t+k} \quad (3.1)$$

où  $x_t$  est la mesure brute qu'on cherche à filtrer,  $y_t$  est le résultat de filtrage, la largeur de la fenêtre de filtrage est  $2m + 1$  et  $\alpha_k$  sont les coefficients de pondération qui vérifient la

relation :

$$\sum_{k=-m}^m \alpha_k = 1.$$

### 3.2.3 Filtre à régression linéaire sur fenêtre mobile (RL)

Comme son nom l'indique, ce filtre permet d'approximer par une droite, la tendance de  $(2m + 1)$  points de mesures successives représentant la largeur de la fenêtre;  $m$  est un entier positif choisi a priori. Ainsi, pour une estimation de la mesure à l'instant  $t$ , la tendance sur cette fenêtre est un segment de droite d'équation :

$$\text{Tendance}(k) = \alpha k + \beta \quad \text{où } -m \leq k \leq m \quad (3.2)$$

$\alpha$  et  $\beta$  sont déterminés par la méthode des moindres carrés. Cette dernière minimise la somme quadratique suivante :

$$S(\alpha, \beta) = \sum_{k=-m}^m (\alpha k + \beta - x_k)^2 \quad (3.3)$$

### 3.2.4 Filtre à régression quadratique (RQ)

Il s'agit ici d'approximer la tendance sur les  $(2m + 1)$  points représentant la largeur de la fenêtre mobile ( $m$ , un entier positif choisi aussi a priori) par un polynôme du second degré :

$$\text{Tendance}(k) = \alpha + \beta k + \gamma k^2 \text{ où } -m \leq k \leq m \quad (3.4)$$

$\alpha$ ,  $\beta$  et  $\gamma$  sont déterminés par la méthode des moindres carrés qui minimise la somme quadratique :

$$S(\alpha, \beta, \gamma) = \sum_{k=-m}^m (\alpha k + \beta k + \gamma k^2 - x_k)^2 \quad (3.5)$$

L'estimation qui nous intéresse est obtenue pour  $(k = 0)$  dans l'équation (3.4). Elle représente aussi la valeur centrale  $\alpha$  de la parabole.

$$\alpha \text{ vérifie : } \frac{\partial S}{\partial \alpha} = 0, \text{ soit, } \gamma \sum_{k=-m}^{k=m} k^2 + (2m+1) \times \alpha = \sum_{k=-m}^{k=m} X(t+k)$$

$$\text{Et } \gamma \text{ vérifie aussi : } \frac{\partial S}{\partial \gamma} = 0, \text{ soit, } \gamma \sum_{k=-m}^{k=m} k^4 + \alpha \sum_{k=-m}^{k=m} k^2 = \sum_{k=-m}^{k=m} k^2 X(t+k)$$

En éliminant  $\gamma$  de ces deux équations, on peut facilement montrer que l'estimation  $\hat{X}_t$  de la donnée  $X_t$  par ce filtre est :

$$\hat{X}_t = \alpha = \frac{\sum_{k=-m}^m X_{t+k} \sum_{k=-m}^m k^4 - \left( \sum_{k=-m}^m k^2 X_{t+k} \right) \left( \sum_{k=-m}^m k^2 \right)}{(2m+1) \sum_{k=-m}^m k^4 - \left( \sum_{k=-m}^m k^2 \right) \left( \sum_{k=-m}^m k^2 \right)} \quad (3.6)$$

### 3.2.5 Filtre médian sur fenêtre mobile (FM)

Ce filtre remédie au problème causé par l'existence de valeurs aberrantes trop bruitées. Il utilise la médiane et non la moyenne pour estimer le paramètre central d'une population. Sur une fenêtre de largeur  $(2m + 1)$  ( $m$  étant un entier positif choisi à priori), le filtre est défini comme suit :

Contrairement aux deux méthodes précédentes, le FM n'est pas paramétrique. Il s'agit tout simplement de prendre la médiane des  $2m + 1$  valeur au voisinage de la valeurs à lisser.

Rappelons d'abord la définition de la médiane d'une série continue  $\{x_1, x_2, \dots, x_n\}$  à partir de l'échantillon ordonné  $\{x_1 < x_2 < \dots < x_n\}$  :

$$Md\{x_1, x_2, \dots, x_n\} = \begin{cases} x_p & \text{si } n=2p+1 \\ \frac{1}{2} (x_p + x_{p+1}) & \text{si } n=2p \end{cases} \quad (3.7)$$

Le filtre est défini comme suit :

$$y_t = Md(x_{t-m}, \dots, x_t, x_{t+m}) \quad (3.8)$$



### 3.2.6 Filtre ébarbeur winsoriseur (WIN)

Ce filtre constitue un compromis entre le filtre médian et le filtre à régression linéaire. La méthodologie employée pour ce filtre est la suivante :

Supposons qu'on ait à estimer la mesure  $X_i$  pour l'instant  $i$ . on prend une fenêtre de largeur  $(2m + 1)$  ( $m$  est un entier positif choisi a priori) centrée en  $X_i$  :

$$X = (X_{i-m}, X_{i-m+1}, \dots, X_i, X_{i+1}, \dots, X_{i+m}).$$

Puis on suit les étapes suivantes :

- a) Classer l'échantillon  $X$  par ordre croissant. L'échantillon obtenu est donc :

$$Y = (Y_{i-m}, Y_{i-m+1}, \dots, Y_i, Y_{i+1}, \dots, Y_{i+m}) \text{ ou } Y_i < Y_j \quad \forall i < j.$$

- b) Éliminer des valeurs trop aberrantes. Il s'agit ici d'une opération d'ébarbage qui exclut un certain nombre de valeurs extrêmes de l'échantillon  $Y$ . Ce rejet peut se faire suivant de deux façons suivantes :

- (1) Rejet a priori : Le nombre de points rejetés est fixé à l'avance. Si, par exemple, on décide d'écarter systématiquement de l'échantillon un certain pourcentage  $a$  (du nombre total  $2m + 1$ ) de dernières valeurs, les nombres des valeurs exclus seront :  $r = \text{Ent}[(2m + 1)a]$  pour les plus faibles valeurs de l'échantillon et  $s = \text{Ent}[(2m + 1)b]$  pour les plus fortes valeurs de l'échantillon.  $\text{Ent}(x)$  représente ici la partie entière du réel  $x$ . Pour une fenêtre mobile, les proportions  $a$  et  $b$  sont prises égales (filtre symétrique) s'il n'y a pas de direction privilégiée dans la contamination. Le filtre élabreur est celui qui estime l'observation  $X_i$  à partir de l'échantillon restant  $(Y_{i-m+r}, Y_i, Y_{i+1}, \dots, Y_{i+m-s})$  comme suit :

$$\hat{X}_i = \frac{1}{(1-a-b)(2m+1)} \times \left[ (1-f_1)Y_{i-m+r} + \sum_{j=-m+r+1}^{j=m-s-1} Y_{i+j} + (1-f_2)Y_{i+m-s} \right] \text{ où (3.9)}$$

$f_1=(2m+1)a-r$  et  $f_2=(2m+1)b-s$  avec  $f_1$  et  $f_2 \in [0;1[$ .

(2) Rejet a postériori : Le nombre de points rejetés est choisi a postériori.

Les valeurs des paramètres  $a$  et  $b$  peuvent être calculées de façon adaptative. On peut, par exemple, rejeter les points qui s'écartent de la moyenne de plus de  $k$  fois l'écart type, calculé sur l'échantillon. D'après Ragot et al. (1990), cette technique possède l'avantage d'écarter les valeurs aberrantes et de ne pas en tenir compte dans les relations de pondération.

c) Réaliser l'opération du Winsorising consiste à substituer les valeurs aberrantes par celles qui leur sont proches. Le filtre winsoriseur est donc celui qui estime l'observation  $X_i$  par la moyenne arithmétique du nouvel échantillon :

$$\underbrace{Y_{i-m+r}, \dots, Y_{i-m+r}}_{(r+1) \text{ termes}}, Y_i, Y_{i+1}, \dots, \underbrace{Y_{i+m-s}, \dots, Y_{i+m-s}}_{(s+1) \text{ termes}}$$

Comme suit :

$$\hat{X}_i = \frac{1}{(2m+1)} \times \left[ r Y_{i-m+r} + \sum_{j=-m+r+1}^{j=m-s-1} Y_{i+j} + s Y_{i+m-s} \right] \quad (3.10)$$

Dorénavant, on notera par «WINp-q», le filtre Winsoriseur, où p représente le nombre de données contenu dans la fenêtre et q le nombre de valeurs, supposées aberrantes, à rejeter.

### 3.2.7 Filtre fréquentiel (FREQ)

Soit un signal discret, périodique de période  $N\delta T$  où  $\delta T$  représente la période d'échantillonnage de la série. Son spectre en fréquence  $X(k)$  est aussi périodique de période  $N\delta f$  échantillonné selon un pas de  $\delta f = 1/N\delta T$  sur un intervalle  $[-f_{\max}; +f_{\max}]$  :

$$X(k\delta f) = \sum_{n=0}^{N-1} x(n) \exp(-j2\pi(k\delta f)(n\delta f)). \quad (3.11)$$

Que l'on peut aussi écrire :

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(-j2\pi \frac{kn}{N}\right). \quad (3.12)$$

Pendant la période d'échantillonnage  $\delta T$  du signal  $\{x(n)\}$  la plus grande fréquence ne doit pas dépasser la fréquence de Nyquist  $f_{\text{Nyquist}} = 1/2\delta T$ . Par conséquent :

$$f_{\max} \leq 1/2\delta T. \text{ Si on note } f_{\max} = k_{\max} \times \delta f \text{ alors : } k_{\max} \leq N/2.$$

Comme on le remarque, la représentation spectrale nécessite un très grand nombre de calculs. Pour les réduire, on fait appel à la transformation de fourrier rapide (T.F.R).

### 3.2.8 Lisseur de Fraser (FRAS)

Dans un environnement univarié où l'on dispose d'une seule série chronologique de mesure, un modèle autorégressif de type AR(p) (Box et Jenkins 1976) est approprié. L'ordre autorégressif « p » doit être fixé à l'avance si on veut mettre au point une technique automatique de lissage des données. En fait, nous nous sommes limités

arbitrairement à l'ordre deux. Ceci est justifié en raison de l'intérêt d'opter pour un modèle parcimonieux (Abraham et Ledolter, 1983). De plus, il est préférable de prendre un modèle autorégressif supérieur à un en raison de l'existence des corrélations entre les mesures effectuées à un à l'instant  $t_n$  et celles effectuées à l'instant  $t_{n-2}$ . D'autre part, en faisant varier l'ordre du modèle autorégressif, il s'est avéré que l'ordre « p » n'avait aucune influence sur les résultats du traitement. La mesure à l'instant «  $t_n$  » peut donc s'exprimer comme suit :

$$X(t_n) = \sum_{k=1}^p b_k X(t_{n-k}) + e(t_n) \quad (3.13)$$

où  $e(t_n)$  est un bruit, considéré blanc.

$b_k$  sont les coefficients régressifs du modèle choisi.

Dans le domaine de la prévision où le modèle autorégressif a été souvent utilisé, on ne possède que des valeurs antérieures pour prédire la valeur future. Dans le domaine du filtrage des données historiques, pour filtrer la mesure prise à l'instant  $t_n$ , on ne possède aussi bien les valeurs antérieures  $t_{n-1}, t_{n-2}, \dots$  que les valeurs subséquentes  $t_{n+1}, t_{n+2}, \dots$

Le lisseur de Fraser (Radix 1984) réalise une estimation optimale de la mesure  $X(t_n)$  à partir de toutes les mesures, au nombre  $N+1$ , effectuées aux instants  $t_0, t_1, \dots, t_n, t_{n+1}, \dots, t_N$ . Il suit ainsi les trois étapes suivantes:

- a. Un filtrage de Kalman «Aller» en partant de  $t_0$  jusqu'à  $t_n$  à partir des conditions initiales  $P(t_0)$  et  $X(t_0)$  permettant d'estimer  $\hat{X}^A(t_n)$  et  $\hat{P}^A(t_n)$ .  $P(t_0)$  étant la matrice de covariance de l'erreur d'estimation du vecteur d'état  $X(t_0)$ .

- b. Un filtrage de Kalman «Retour» en partant de  $t_0$  jusqu'à  $t_n$  à partir des conditions initiales  $P(t_N)$  et  $X(t_N)$  permettant d'estimer  $\hat{X}^R(t_n)$  et  $\hat{P}^R(t_n)$ .  $P(t_N)$  étant la matrice de covariance retour de l'erreur d'estimation du vecteur d'état  $X(t_N)$ .
- c. Une combinaison optimale des estimations aller  $\hat{X}^A(t_n)$  et retour  $\hat{X}^R(t_n)$  à partir des relations suivantes :

$$\begin{cases} \hat{P}^{-1}(t_n) = [\hat{P}^A(t_n)]^{-1} + [\hat{P}^R(t_n)]^{-1} \\ \hat{X}(t_n) = \hat{P}(t_n) \{ [\hat{P}^A(t_n)]^{-1} \hat{X}^A(t_n) + [\hat{P}^R(t_n)]^{-1} \hat{X}^R(t_n) \} \end{cases} \quad (3.14)$$

L'estimation optimale ou la plus probable est donc une combinaison linéaire des deux estimations «aller » et «retour » qui favorise toujours celle dont la «variance» est la plus faible. Les coefficients de pondération sont représentés par des matrices de précision  $(P^A)^{-1}$  et  $(P^R)^{-1}$ . Il faut cependant souligner un détail important dans l'application du système d'équations (3.14). Le filtre de Kalman, appliqué selon la technique MISP (Todini 1978), fournit à chaque pas de temps, dans les sens «Aller » et «Retour », deux estimations du vecteur d'état :  $\hat{X}^A(t_n/t_{n-1})$  et  $\hat{X}^A(t_n/t_n)$  à l'aller,  $\hat{X}^R(t_n/t_{n-1})$  et  $\hat{X}^R(t_n/t_n)$  au retour.  $\hat{X}^A(t_n/t_{n-1})$  est la valeur de  $X$  prédite, dans le sens « Aller », à partir des valeurs mesurées de  $t_0$  à  $t_{n-1}$ .  $\hat{X}^A(t_n/t_n)$  est la valeur de  $X$  estimée à l'« Aller » après la réception de la mesure à l'instant  $t_n$ . la notation est similaire pour  $X^R$ . afin de ne pas utiliser la mesure à l'instant  $t_n$  deux fois , il faut utiliser la valeur prévue dans un sens , par exemple  $\hat{X}^A(t_n/t_{n-1})$  et la valeur filtrée dans l'autre sens  $\hat{X}^R(t_n/t_n)$ . Dorénavant, on notera par «FRAS », le filtre Fraser.

Pour remédier aux problèmes d'autocorrélation des résidus, lors de l'utilisation de la méthode standard des moindres carrés, l'identification de ces paramètres s'effectue par application du filtre récursif et adaptatif de Kalman, (Bennis, 1987).

### 3.3 Mise en œuvre des techniques de filtrage : Données synthétiques

#### 3.3.1 Introduction

Avant d'appliquer les techniques de filtrage décrites précédemment sur des données réelles, il nous a semblé intéressant de les appliquer. L'intérêt d'une telle démarche est de valider notre méthodologie. Il s'agit d'un hydrogramme et d'un pollutogramme générés à partir d'une pluie synthétique simulée sur un bassin urbain. Pour rendre les données plus proches de la réalité, nous leur avons rajouté, un bruit de distribution normale et d'écart type constant. Nous avons ensuite modifié, volontairement et de façon importante, un certain nombre de valeurs de ces deux courbes synthétiques (figures, 3.1 à 3.16).

#### 3.3.2 Description des données synthétiques

Pour générer l'hydrogramme synthétique, nous avons utilisé un bassin de drainage réel, situé dans la Ville de Verdun, au Québec. Les caractéristiques physiques de ce bassin ont été utilisées pour générer tout d'abord un hydrogramme de crue à partir d'un événement pluviométrique synthétique. Pour cela, nous avons utilisé le modèle du réservoir non linéaire du logiciel SWMM (Huber et Dickinson, 1988). L'hydrogramme synthétique utilisé est obtenu enfin, par utilisation du modèle d'onde dynamique, représenté par les équations de Saint Venant. On rappelle que ces équations tiennent compte de tous les phénomènes physiques importants de transfert d'onde.

Pour générer le pollutogramme synthétique, on a utilisé un modèle simple de «rating curve» (Huber et Dickson, 1988) que nous avons modifié pour tenir compte d'un effet de premier lavage éventuel (Bennis et al., 2001). Il est exprimé comme suit :

$$CH(t) = a(Q(t - n\delta t))^b \quad (3.15)$$

$CH(t)$  est la charge de MES à l'instant  $t$ .  $n\delta t$  est un décalage de temps que nous avons volontairement rajouté pour tenir compte du retard ou de l'avance entre la pointe de l'hydrogramme et celle du pollutogramme, observé dans certains bassins urbains. Lorsqu'il y a synchronisme entre l'hydrogramme et le pollutogramme,  $n = 0$ .  $\delta t$ , étant l'intervalle de calcul pris en considération pour générer l'hydrogramme précédent,  $a$  et  $b$  sont des constantes à déterminer.

### 3.3.3 Critères de performances

#### 3.3.3.1 Efficacité

Pour apprécier la performance des techniques de filtrage utilisées, on introduit l'indice suivant :

$$\eta = 1 - \frac{\varepsilon_{est}^2}{\varepsilon_b^2} \quad (3.16)$$

$\varepsilon_{est}^2$  est la moyenne des écarts quadratiques entre les données estimées ( $x_{est}$ ) par une technique de filtrage et les données initiales ( $x$ ) non bruitées.  $\varepsilon_b^2$  est la moyenne des écarts quadratiques entre les données initiales bruitées ( $x_b$ ) et les données initiales idéales :

$$\begin{cases} \varepsilon_{est}^2 = \frac{1}{N} \sum_{j=1}^N (x_{est}(j) - x(j))^2 \\ \varepsilon_b^2 = \frac{1}{N} \sum_{j=1}^N (x_b(j) - x(j))^2 \end{cases} \quad (3.17)$$

On voit donc que la technique de filtrage est efficace si l'indice  $\eta$  est positif. La performance d'une technique se mesure par rapprochement de  $\eta$  à l'unité (1).

### 3.3.3.2 Coefficient des performance

Pour comparer les techniques entre elles, on introduit le paramètre P de performance :

$$P = \frac{\eta - \eta_{\min}}{1 - \eta_{\min}} \times 100 \quad (3.18)$$

$\eta_{\min}$  est l'efficacité minimale obtenue parmi toutes les techniques utilisées. P exprime donc l'amélioration apportée par un filtre donné, par rapport au filtre le moins efficace.

### 3.3.3.3 Autocorrélation des résidus

Le deuxième critère de comparaison porte sur les coefficients d'autocorrélation des résidus. Les méthodes et solutions proposées sont basées sur l'hypothèse que les résidus prennent des valeurs aléatoires de moyenne nulle et de variance constante. Il faut par conséquent vérifier que leurs coefficients d'autocorrélation s'approchent de zéro.

Une estimation du coefficient autorégressif d'ordre k peut être définie par :

$$\hat{\rho}_k = \frac{\sum_{t=1}^{N-k} [\text{Res}(t) - \overline{\text{Res}}] [\text{Res}(t+k) - \overline{\text{Res}}]}{\sum_{t=1}^k [\text{Res}(t) - \overline{\text{Res}}]^2} \quad (3.19)$$

$\text{Res}(t) = x_b(t) - x_{\text{est}}(t)$ ,  $\overline{\text{Res}}$  est la moyenne des résidus et  $x_b$  est le signal qu'on veut filtrer.

Grâce à l'approximation de Barlett (Box et Jenkins, 1976), l'écart type des estimations s'écrit :



$$S(r_k) = \left( 1 + 2 \sum_{j=1}^{k-1} r_j^2 \right)^{1/2} N^{-1/2} ; \quad k = 1, 2, 3 \dots \quad (3.20)$$

N est le nombre total (ou un échantillon) de données.

Pour ne pas rejeter l'hypothèse  $H_0$  ( $H_0: \rho_k=0; k=1,2,3\dots$ ) de la nullité théorique du coefficient d'autocorrélation des résidus, on utilise le test de Student défini par (Pankratz 1983) :

$$t_k = \frac{r_k - 0}{S(r_k)} ; \quad k = 1, 2, 3 \dots \quad (3.21)$$

En pratique, si la valeur absolue de  $t_k$  est inférieure à 1.25, pour  $k = 1, 2, 3$ , et inférieure à 1.6 pour  $k > 3$ , l'hypothèse  $H_0$  est vraie et les résidus seraient alors considérés indépendants.

### 3.3.4 Résultats

Les résultats de la validation du pollutogramme et de l'hydrogramme synthétiques, perturbés artificiellement par un bruit de petites et grandes amplitudes, sont donnés dans les figures 3.1 à 3.16. Les courbes représentent un même échantillon de 40 données espacées par un intervalle de 5 minutes. Les représentations montrent les résultats de l'application de chacun des filtres exposés précédemment au pollutogramme et à l'hydrogramme synthétiques.

L'inspection visuelle de ces graphiques révèle clairement que le filtre à régression linéaire est le moins performant parmi tous les filtres utilisés. On constate ainsi sur la figure 3.3 que les valeurs aberrantes introduites ont contaminé les valeurs voisines dans le processus de filtrage. Ceci n'est pas le cas pour le filtre médian (fig. 3.1), winsoriseur (fig. 3.5 et 3.7) qui ont l'avantage de ne pas utiliser directement les valeurs suspectes dans le calcul de la valeur filtrée. Les conclusions sont similaires pour les hydrogrammes montrés sur les graphiques (3.9), (3.11), (3.13), et (3.15). Le tableau

(3.1) qui chiffre l'efficacité et la performance de chacun des filtres confirme ces constatations. Ainsi, dans le cas du pollutogramme, le filtre qui a la plus faible efficacité est le filtre à régression linéaire RL7 pour lequel  $\eta = 0.90$ . La performance des trois autres filtres a été calculée relativement à ce dernier. Elle vaut 86.56% pour le filtre winsoriseur WIN7-4, 85.4% pour le filtre WIN5-2 et 83.67% pour le filtre FM5 lorsque appliqués sur les pollutogrammes bruités. Ces chiffres doivent être interprétés au même titre qu'un coefficient de détermination. À titre d'exemple, le fait que  $P = 86.56\%$  pour le filtre WIN7-4 signifie que ce dernier a réussi de récupérer 86.56% de l'information que le filtre de base, en l'occurrence RL7, n'a pas réussi de retrouver

Chacun des filtres présentés précédemment possède des avantages et des inconvénients. Afin d'améliorer les résultats obtenus précédemment, nous avons exploré la possibilité d'appliquer successivement plusieurs filtres à la même série initiale bruitée. Compte tenu du nombre de filtres disponibles, le nombre de permutations possibles est imposant.

Après plusieurs tests, quatre combinaisons ont été retenues pour comparaison. Il s'agit d'appliquer successivement un filtre de Fraser, puis un filtre fréquentiel aux valeurs filtrées par chacun des quatre filtres précédents. La comparaison portera par conséquent sur les quatre scénarios suivants :

- 1- RL7 + FRAS + FRÉQ
- 2- M5 + FRAS + FRÉQ
- 3- WIN5-2 + FRAS + FRÉQ
- 4- WIN7-4 + FRAS + FRÉQ

Le tableau (3.2) indique que le filtrage à l'aide de WIN5-2 suivi d'un filtre de Fraser puis d'un filtre fréquentiel donne la meilleure performance. L'utilisation d'un filtre WIN7-4 ou FM5 donne une performance légèrement inférieure. Par contre, l'utilisation

des filtres de Fraser et fréquentiel n'ont pas réussi à améliorer les résultats médiocres obtenus par le filtre RL7.

Selon les tableaux (3.3a) et (3.3b) qui portent sur le test de nullité des coefficients d'autocorrélation des résidus pour le pollutogramme, toutes les valeurs du coefficient de student t pour les trois premiers retards sont inférieures à 1,25. Les tableaux (3.3c) et (3.3d) indiquent des résultats similaires pour l'hydrogramme.

Compte tenu de ces résultats, on peut affirmer que l'application successive de plusieurs méthodes de filtrage donne des résultats meilleurs que l'application isolée de chacun des filtres. Les deux étapes de filtrage qui suivent la première étape permettent de corriger «au mieux», les imperfections obtenues par le premier filtrage. Les scénarios (3) et (4) donnent la meilleure performance respectivement pour le pollutogramme et l'hydrogramme. Il est important de souligner aussi, qu'il est nécessaire de voir la représentation fréquentielle du signal obtenu à l'étape 2 afin de choisir la fréquence de coupure la plus appropriée. En effet, une sur ou sous-estimation de la fréquence de coupure entraîne généralement une sous ou surestimation des pointes. Une fréquence de coupure  $f_c = 0.2$  s'est révélée optimale dans le cadre de la présente étude. Au même titre, la largeur de la fenêtre des filtres médian et winsoriseur doit être choisie avec précaution. Une largeur de fenêtre trop petite, risque d'inclure certaines valeurs aberrantes dans le cas de plusieurs perturbations successives. Une trop grande largeur de fenêtre fera systématiquement abaisser les pointes des hydrogrammes et pollutogrammes. Dans le cadre de la présente étude, la fenêtre de largeur 5 s'est révélée optimale.

### 3.4 Mise en œuvre sur des données réelles

#### 3.4.1 Description du bassin urbain

L'étude porte sur la validation des données liées à la qualité des eaux usées et pluviales de Ville de Verdun. Le site d'étude est le secteur I, représenté par la figure 3.17. Le bassin est limité à l'Ouest par le canal de l'aqueduc et l'usine de filtration de la ville de Montréal. Au Nord, le secteur est limité par le territoire de la ville de Montréal. Au sud, il est limité par la première avenue et finalement à l'Est par le fleuve Saint-Laurent.

Le réseau d'égout unitaire est constitué en grande partie de conduites de forme ovoïdale (600 x 900) installées selon de faibles pentes et incapables d'acheminer un débit 1/5 ans sans débordement. Ainsi, le réseau unitaire du secteur I de la ville de Verdun a fait l'objet de plusieurs inondations dans plusieurs de ses tronçons. En vue de le restructurer, la municipalité a entrepris la construction d'un nouveau réseau de «subsurface» pour drainer le surplus d'eau pluviale et l'acheminer directement au fleuve Saint-Laurent soulageant ainsi le réseau existant et évitant les débordements fréquents d'eau unitaire dans le fleuve Saint-Laurent.

Les concentrations de polluants ont été mesurées à l'entrée de la station de pompage Rhéaume le 30/09/1999 (figures 3.18, 3.19 et 3.20) et le 13/10/1999 (figures 3.21, 3.22 et 3.23) et la station de Wellington le 30/09/1999 (figures 3.24, 3.25 et 3.26) et le 13/10/1999 (figures 3.27, 3.28 et 3.29). Ces points ont été choisis car actuellement le débit y représente la résultante de toutes les eaux pluviales, sanitaires, industrielles et d'infiltration du secteur I.

### 3.5 Résultats

L'objectif de cette deuxième partie est de filtrer les mesures de concentrations de matières en suspension prises à la station de Rhéaume (figures 3.18, 3.19, 3.20, 3.21, 3.22 et 3.23), lors de l'orage du 30/09/1999 et du 13/10/1999. La station est équipée d'un seul appareil de mesure. Cette contrainte ne permet donc pas d'appliquer des techniques de validation utilisant une redondance matérielle. Par ailleurs, les concentrations de polluants sont reliées à tellement de variables (débit, précipitation, durée de temps sec, heure dans la journée de l'événement pluvial, jour de la semaine...) qu'il est difficile de les regrouper dans un modèle analytique type pouvant servir à valider les données. Par conséquent, il est logique de recourir aux modèles univariés tels que ceux décrits en première partie.

Les figures 3.18, 3.19 et 3.20 présentent les résultats obtenus par trois types validation utilisant le procédé décrit précédemment. L'étape 1 de ce procédé est respectivement, le Win5-2 pour la figure 3.18, le FM5 pour la figure 3.19 et le RL5, pour la figure 3.20. L'étape 2 étant celle du filtre de Fraser et l'étape 3, celle du filtre passe bas, ici de fréquence de coupure égale à 0.3.

Comme on peut le constater sur les trois figures, les trois procédés donnent de bons résultats. Ceci est aussi confirmé par le tableau 3.4 qui montre l'amélioration systématique des coefficients d'autocorrélation apportés par les trois types de filtrage. Néanmoins, on peut voir (figures 3.18, 3.19 et 3.20), que les procédés utilisant en première étape le WIN5-2 et le RL5 (figures 3.18 et 3.19) sont ceux qui se comportent le mieux. En effet, ils vérifient, de façon satisfaisante, le critère de conservation des pointes du pollutogramme. Ils obéissent aussi, au mieux, au critère de conservation de la masse polluante obtenue par le produit cumulé des concentrations et des débits correspondants: 101% pour RL5 et 97% pour WIN5-2, par rapport à 94% pour FM5.

Il est important de dégager à ce niveau, que le procédé utilisant le RL5, en première étape, a donné, ici de meilleurs résultats que dans le cas synthétique. Le contraire s'est passé pour le filtre médian. Ceci prouve, encore une fois, que dans un environnement univarié où il n'existe ni redondance matérielle ni redondance analytique, il est difficile de trouver une méthode universelle pour filtrer les mesures. C'est à cet effet que l'une des raisons pour laquelle nous proposons une variété de techniques de filtrage et un certain nombre de critères de performance pour aider les concernés à filtrer au mieux leurs données. Le bon sens de l'ingénieur, ainsi que sa connaissance du terrain vont être un facteur, souvent déterminant pour une excellente validation de données.

### 3.6 Conclusion

Plusieurs techniques de filtrage ont été appliquées sur des séries chronologiques de débit et de concentration de polluants en milieu urbain. Les résultats ont été comparés à l'aide de critères de performance quantitatifs et qualitatifs. Plusieurs techniques de filtrage donnent des résultats satisfaisants. Cependant, l'application successive de plusieurs méthodes de filtrage donne des résultats meilleurs que l'application isolée de chacune des techniques. Lorsque appliquée sur des hydrogrammes et pollutogrammes dans un réseau d'assainissement, la méthodologie poursuivie donne de bons résultats aussi bien pour les séries synthétiques bruitées que réelles.

Tableau 3.1

Efficacité et performance des filtres utilisés

Filtres	Pollutogramme	
	$\eta$	P(%)
RL7	0.90	0
RL7 + FRAS + FRÉQ	0.90	2.80
FM5	0.98	83.67
FM5 + FRAS + FRÉQ	0.98	85.28
WIN5-2	0.98	85.40
WIN5-2 + FRAS + FRÉQ	0.99	87.67
WIN7-4	0.99	86.56
WIN7-4 + FRAS + FRÉQ	0.99	78.58
$\eta$ = efficacité		P = Performance

Tableau 3.2

Efficacité et performance des filtres utilisés

Filtres	Hydrogramme	
	$\eta$	P(%)
RL7	0.89	0
RL7 + FRAS + FRÉQ	0.90	6.16
FM5	0.92	29.40
M5 + FRAS + FRÉQ	0.95	54.99
WIN5-2	0.94	47.08
WIN5-2 + FRAS + FRÉQ	0.95	53.44
WIN7-4	0.94	43.62
WIN7-4 + FRAS + FRÉQ	0.97	74.89
$\eta$ = efficacité		P = Performance



Tableau 3.3a

Test de nullité des coefficients d'autocorrélation des résidus pour le pollutogramme

Retard	Signal synthétique		Filtre M5		M5 Fras, Fréq 2		Filtre RL7		RL7 Fras, Fréq. 2	
	R	T	R	T	R	T	R	T	R	T
1	-.01	0	-.06	-.40	-.06	-.41	-.11	-.70	-.14	-.86
2	-.07	0	-.12	-.78	-.09	-.60	-.22	-1.35	-.17	-1.06
3	-.12	0	-.13	-.79	-.13	-.80	-.28	-1.65	-.17	-1.02
4	-.15	0	-.10	-.64	-.11	-.70	-.02	-.10	-.13	-.79

Tableau 3.3b

Test de nullité des coefficients d'autocorrélation des résidus pour le pollutogramme

Retard	Filtre		Win5-2		Filtre		Win7-4	
	Win5-2		Fras, freq.2		Win7-4		Fras, freq.2	
	R	T	R	T	R	T	R	T
1	-.10	-.61	-.07	-.44	-.06	-.37	-.03	-.22
2	-.12	-.72	-.11	-.70	-.12	-.76	-.10	-.60
3	-.11	-.66	-.13	-.82	-.14	-.87	-.14	-.89
4	-.08	-.51	-.13	-.77	-.12	-.76	-.18	-1.09

Tableau 3.3c

Test de nullité des coefficients d'autocorrélation des résidus pour l'hydrogramme

Retard	Signal		Filtre		M5		Filtre		RL7	
	synthétique		M5		Fras, freq.2		RL7		Fras, freq.2	
	R	T	R	T	R	T	R	T	R	T
1	.08	0	-.10	-.64	.01	.09	.02	.11	-.08	-.49
2	.01	0	-.03	-.19	-.05	-.31	-.11	-.69	-.05	-.33
3	-.34	0	-.27	-1.71	-.38	-2.42	-.49	-3.06	-.37	-2.32
4	-.35	0	-.28	-1.64	-.37	-2.07	-.25	-1.30	-.39	-2.15

Tableau 3.3d

Test de nullité des coefficients d'autocorrélation des résidus pour l'hydrogramme

Retard	Filtre		Win5-2		Filtre		Win7-4	
	Win5-2		Fras, freq.2		Win7-4		Fras, freq.2	
	R	T	R	T	R	T	R	T
1	-.10	-.61	-.07	-.44	-.06	-.37	-.03	-.22
2	-.12	-.72	-.11	-.70	-.12	-.76	-.10	-.60
3	-.11	-.66	-.13	-.82	-.14	-.87	-.14	-.89

Tableau 3.4

Coefficient d'autocorrélation des concentrations MES

<b>Autocorrélation</b>	$\rho_1$	$\rho_2$	$\rho_3$	$\rho_4$	$\rho_5$
<b>Mesures réelles</b>	0.833	0.746	0.644	0.424	0.243
WIN5-2, FRAS, FREQ.3	0.887	0.801	0.651	0.432	0.215
M5, FRAS, FREQ.3	0.953	0.833	0.661	0.433	0.217
RL5, FRAS, FREQ.3	0.956	0.851	0.706	0.507	0.290

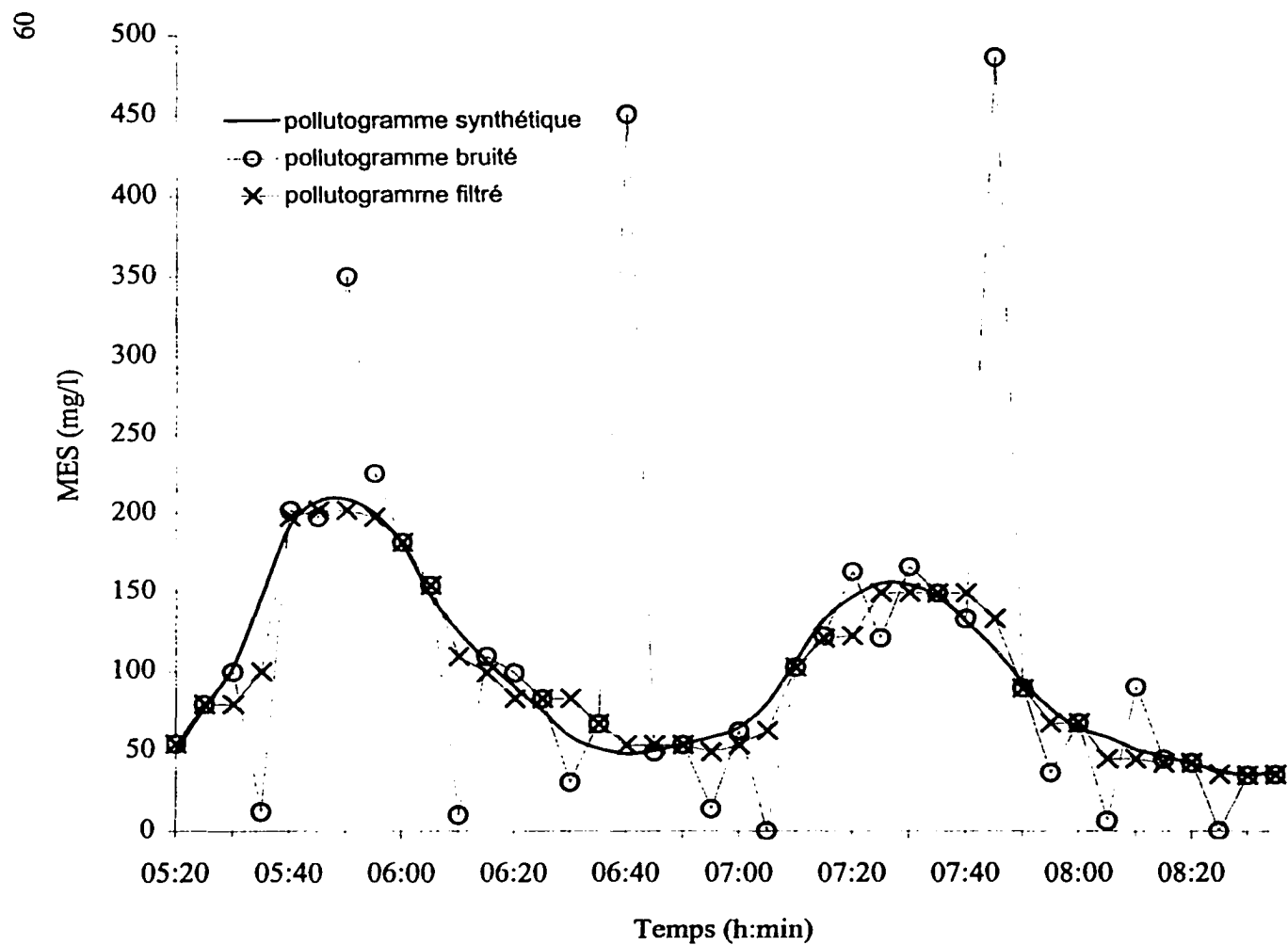


Figure 3.1 Filtrage du pollutogramme bruité- application du filtre médian.

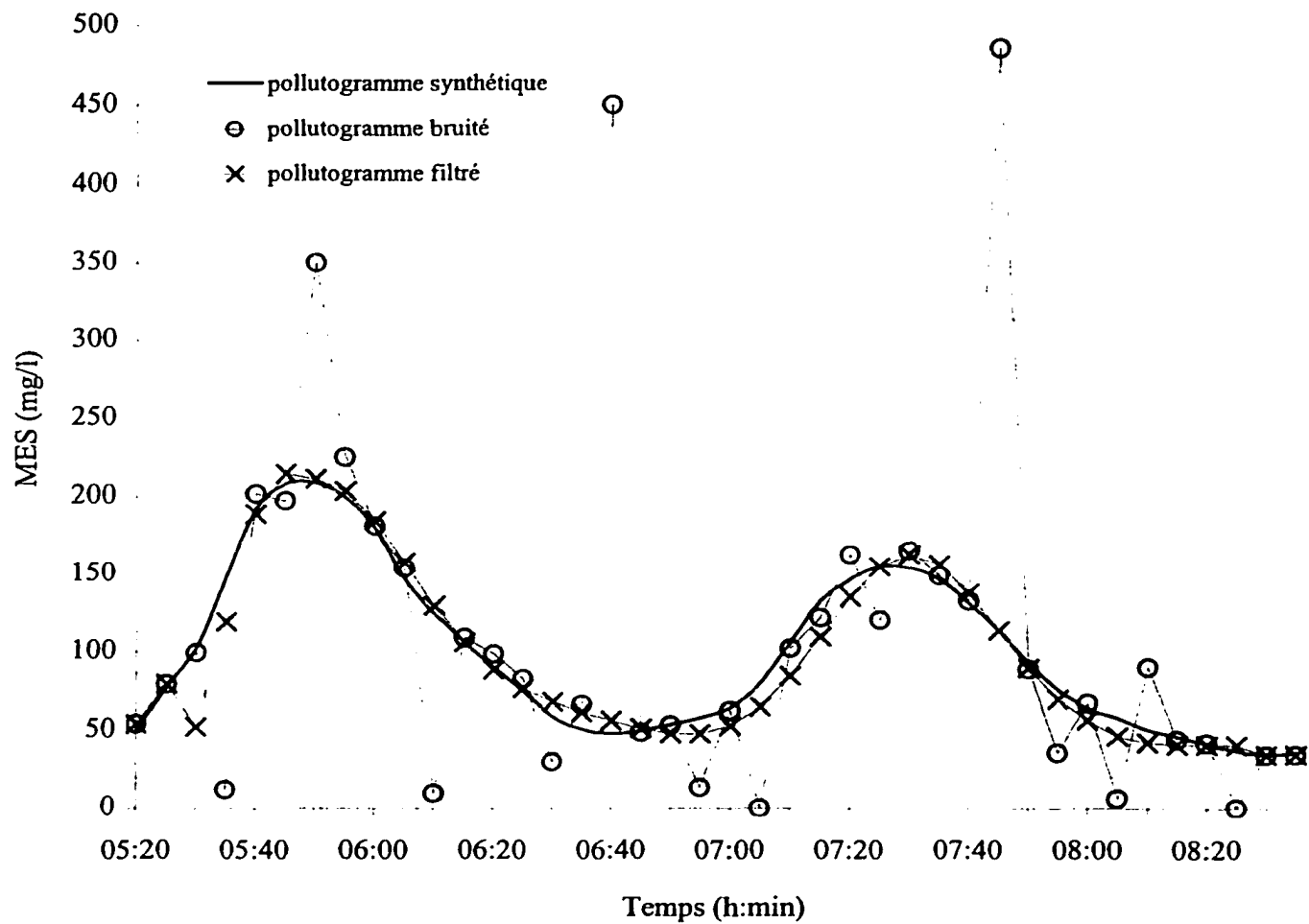


Figure 3.2 Filtrage du pollutogramme bruité- filtre médian 5 puis fraser et fréquentiel 0.2

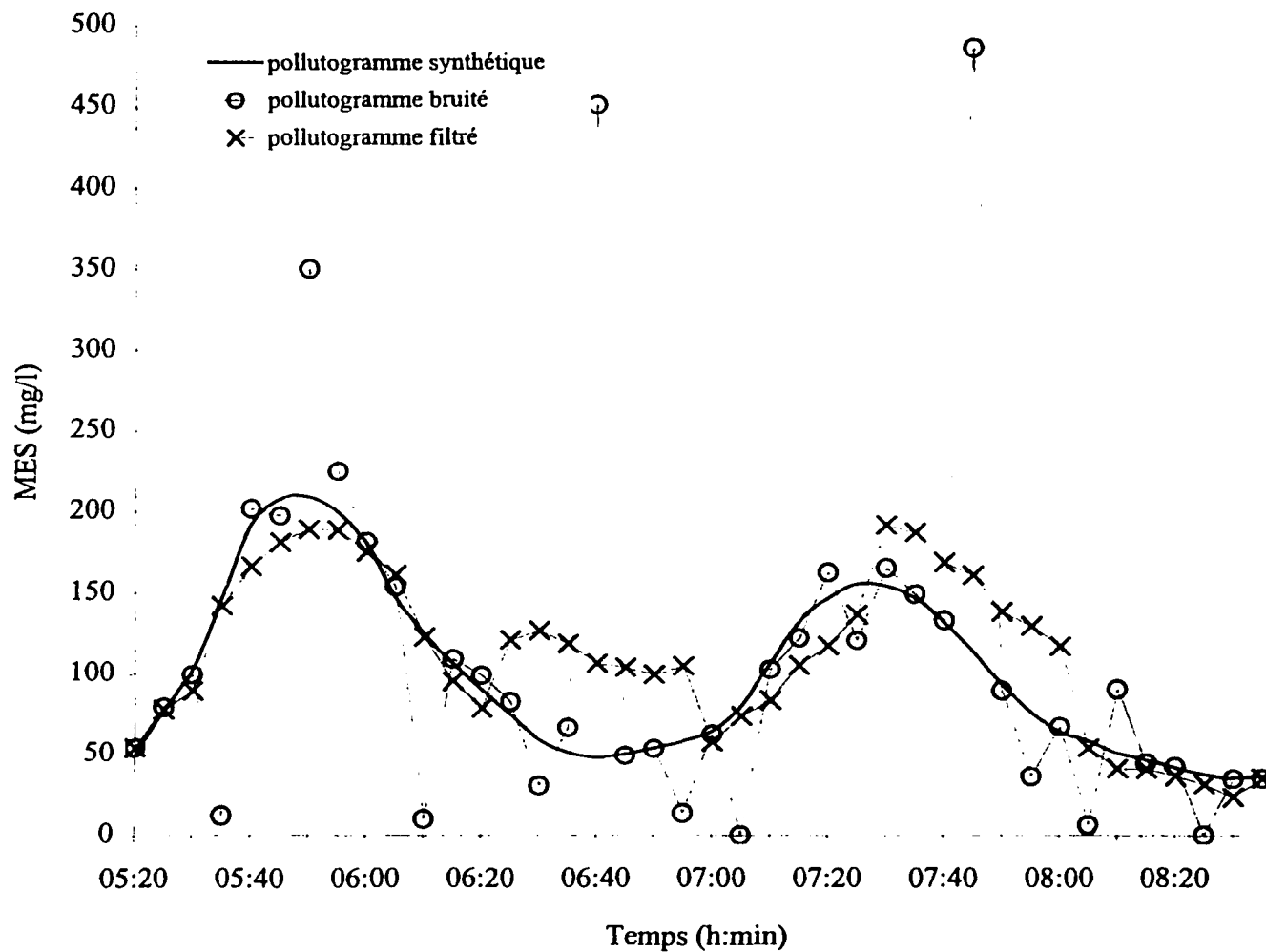


Figure 3.3 Filtrage du pollutogramme bruité- filtre à régression linéaire de paramètre 7



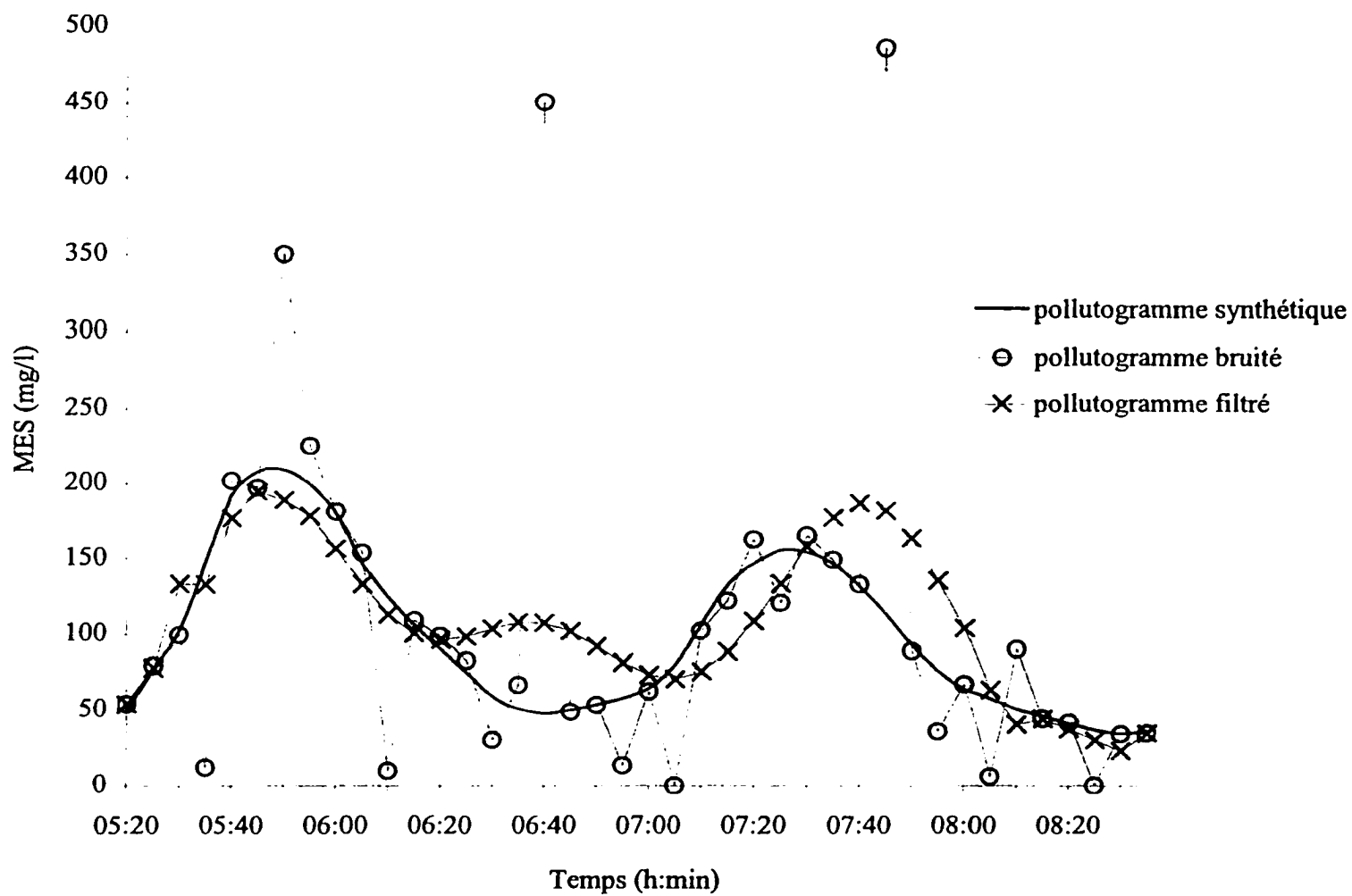


Figure 3.4 Filtrage du pollutogramme bruité- filtre à régression linéaire 7 puis fraser et fréquentiel 0.2

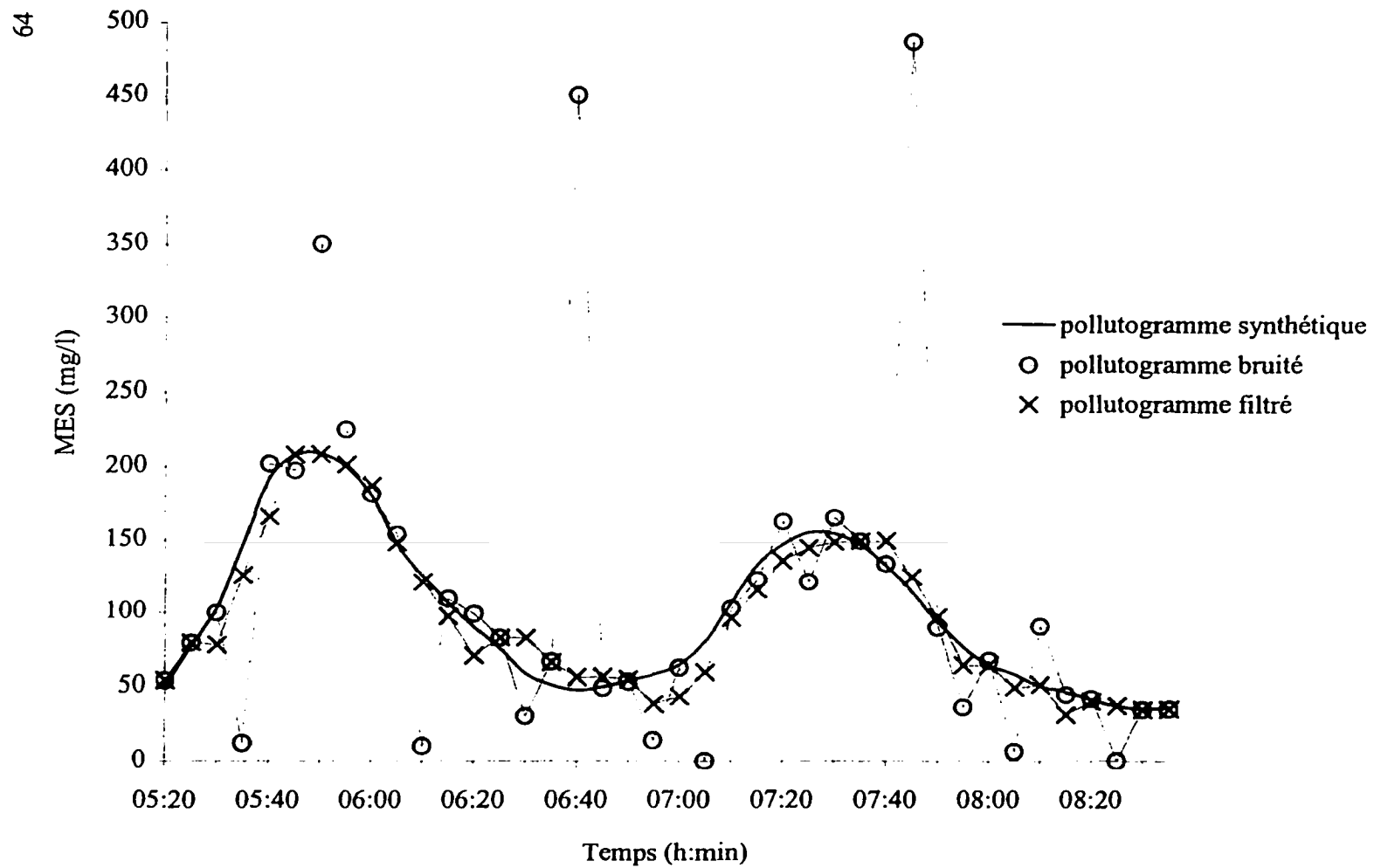


Figure 3.5 Filtrage du pollutogramme bruité- filtre winsoriseur de paramètres 5,2

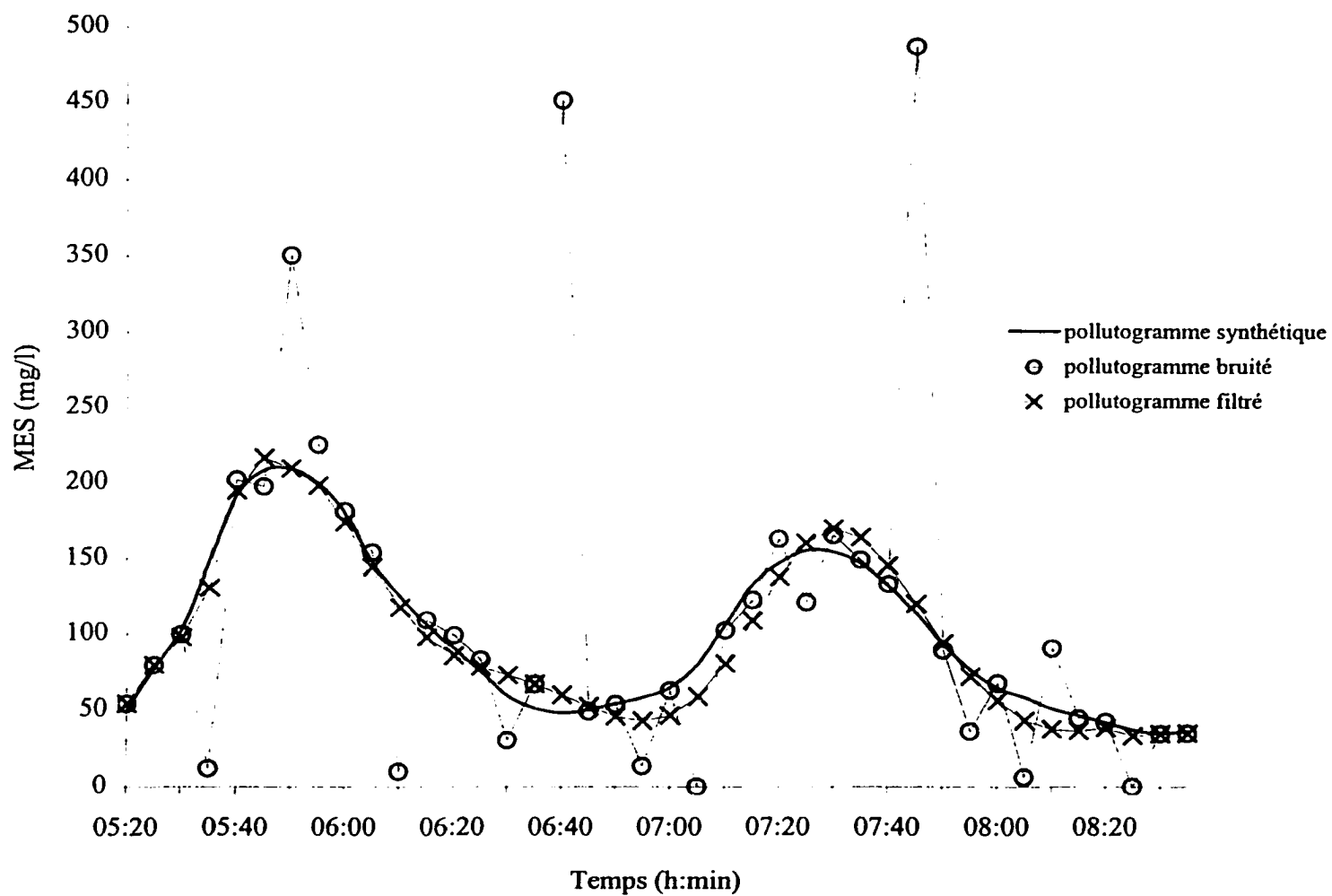


Figure 3.6 Fitrage du pollutogramme bruité- filtre winsoriseur 5,2 puis fraser et fréquentiel 0.2

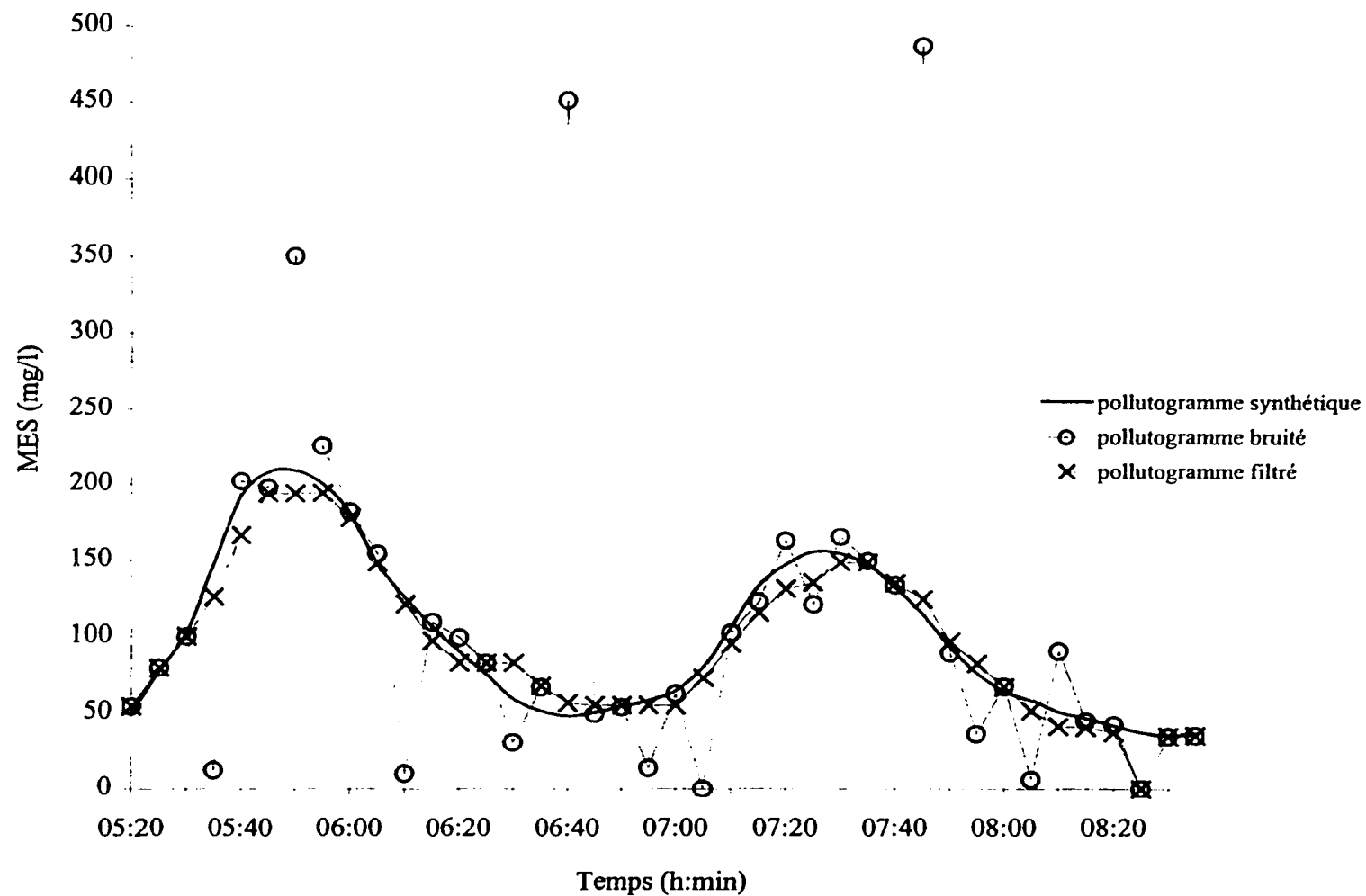


Figure 3.7 Filtrage du pollutogramme bruité- filtre winsoriseur de paramètres 7,4

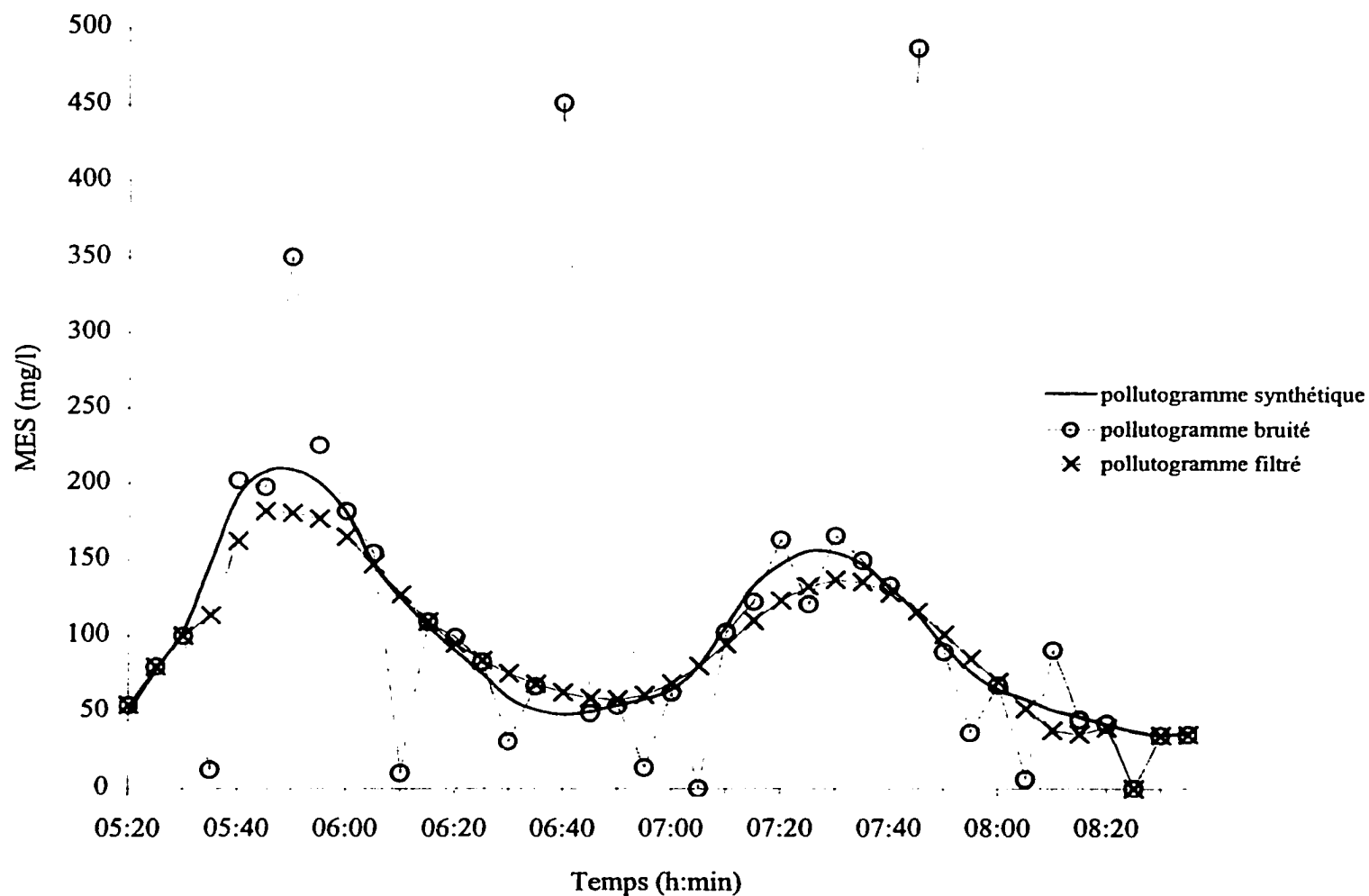


Figure 3.8 Filtrage du pollutogramme bruité- winsoriseur 7,4 puis fraser et fréquentiel

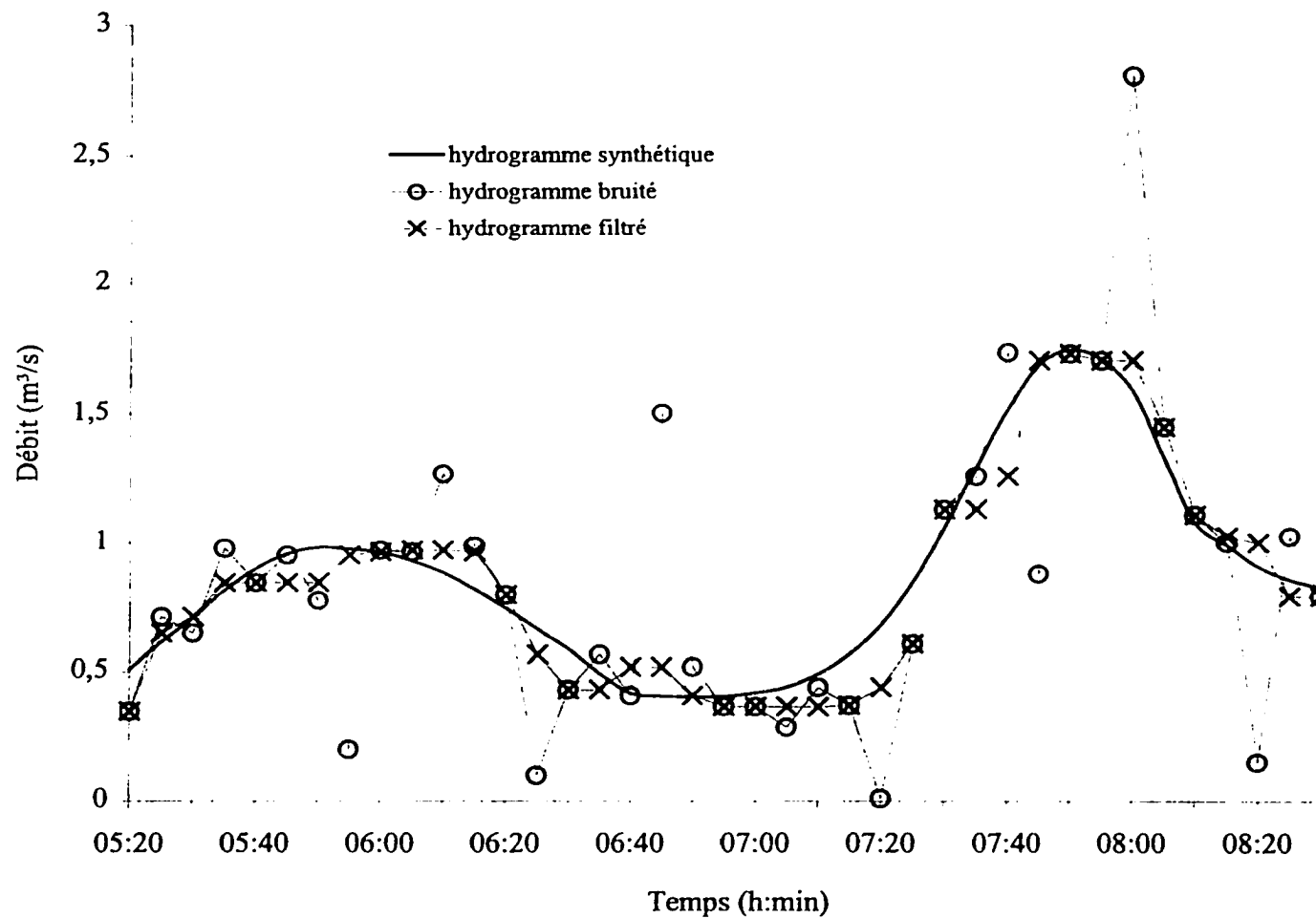


Figure 3.9 Filtrage du hydrogramme bruité- filtre médian de paramètre 5

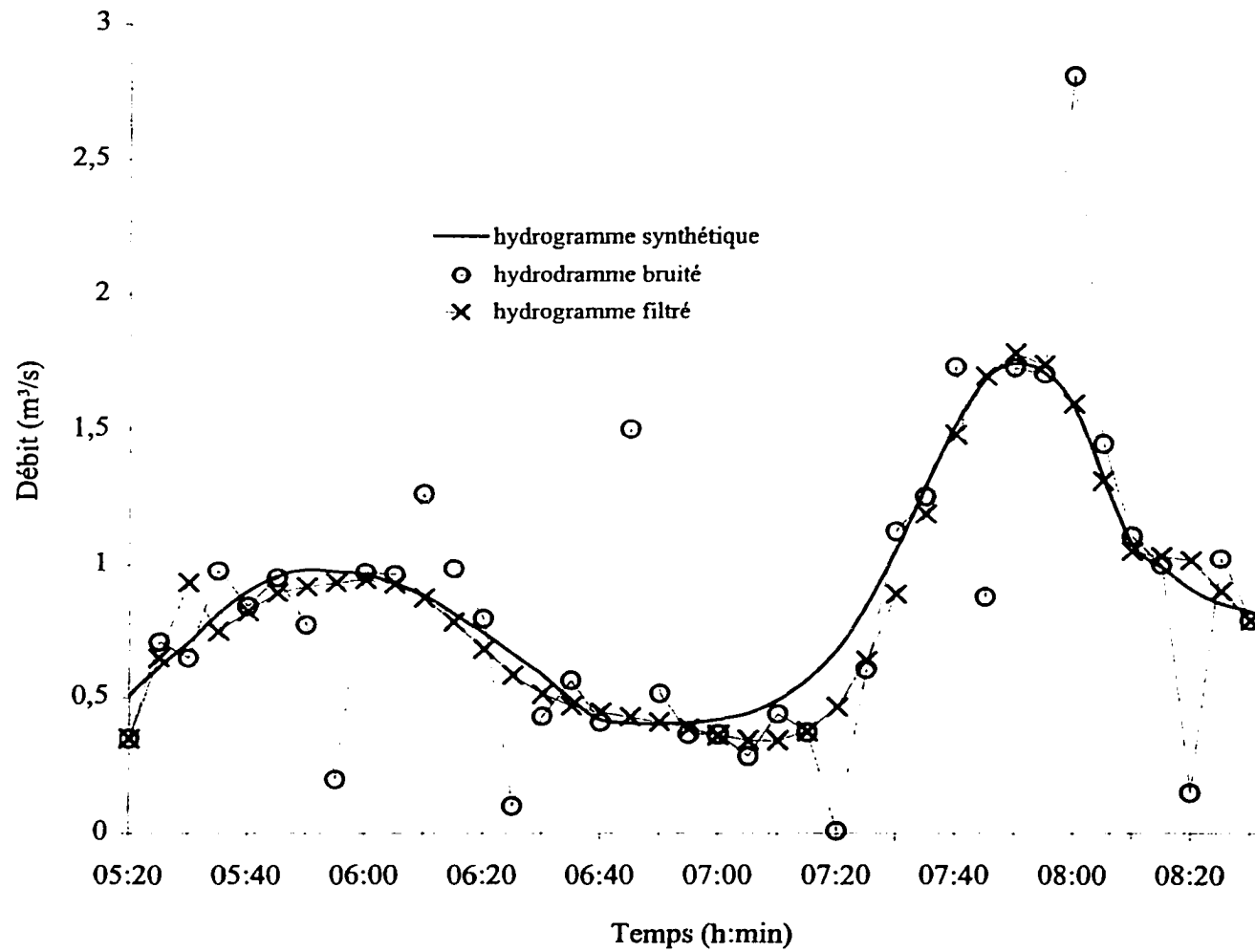


Figure 3.10 Filtrage de l'hydrogramme bruité- médian 5 puis fraser et fréquentiel 0.2

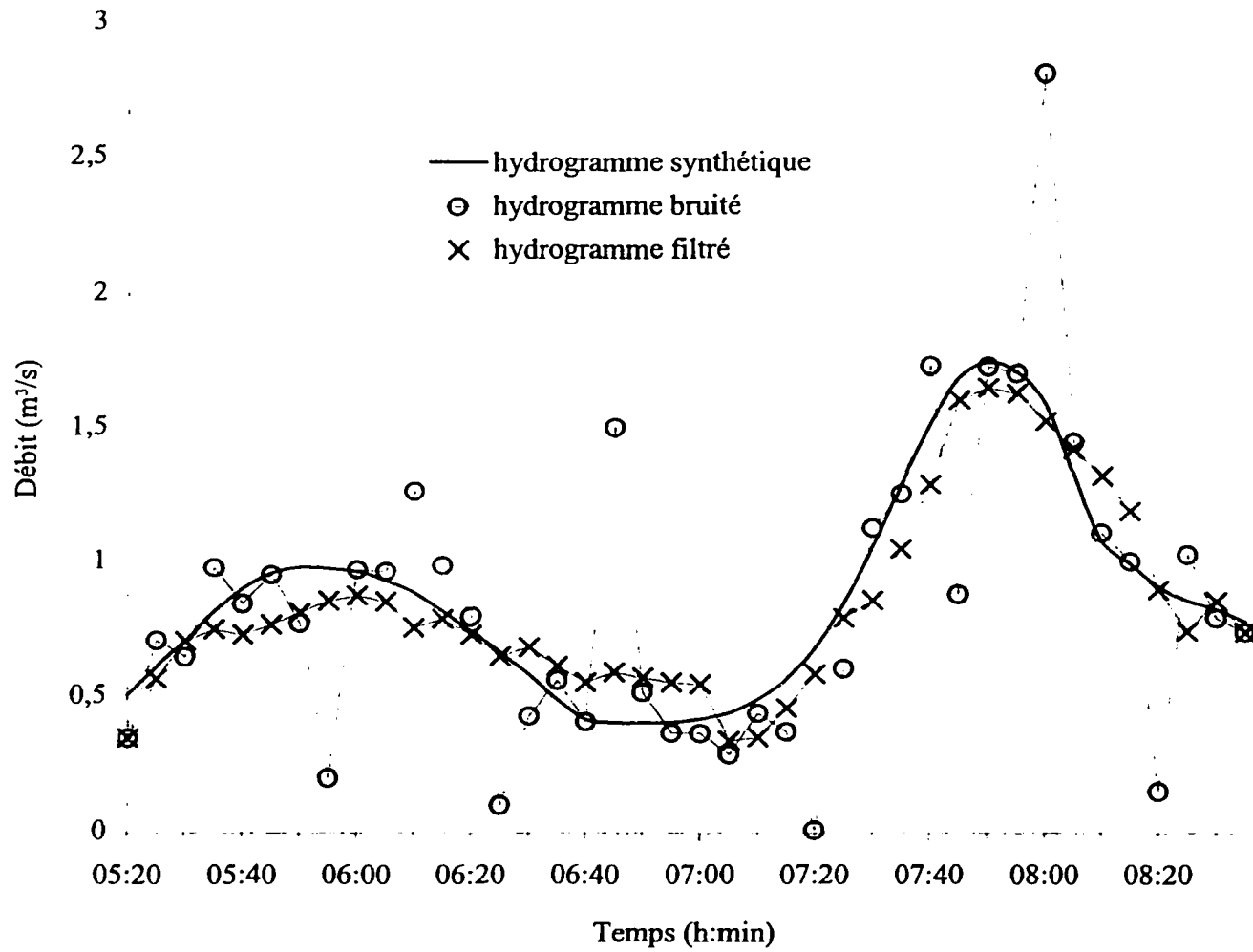


Figure 3.11 Filtrage de l'hydrogramme bruité- filtre à régression linéaire 7



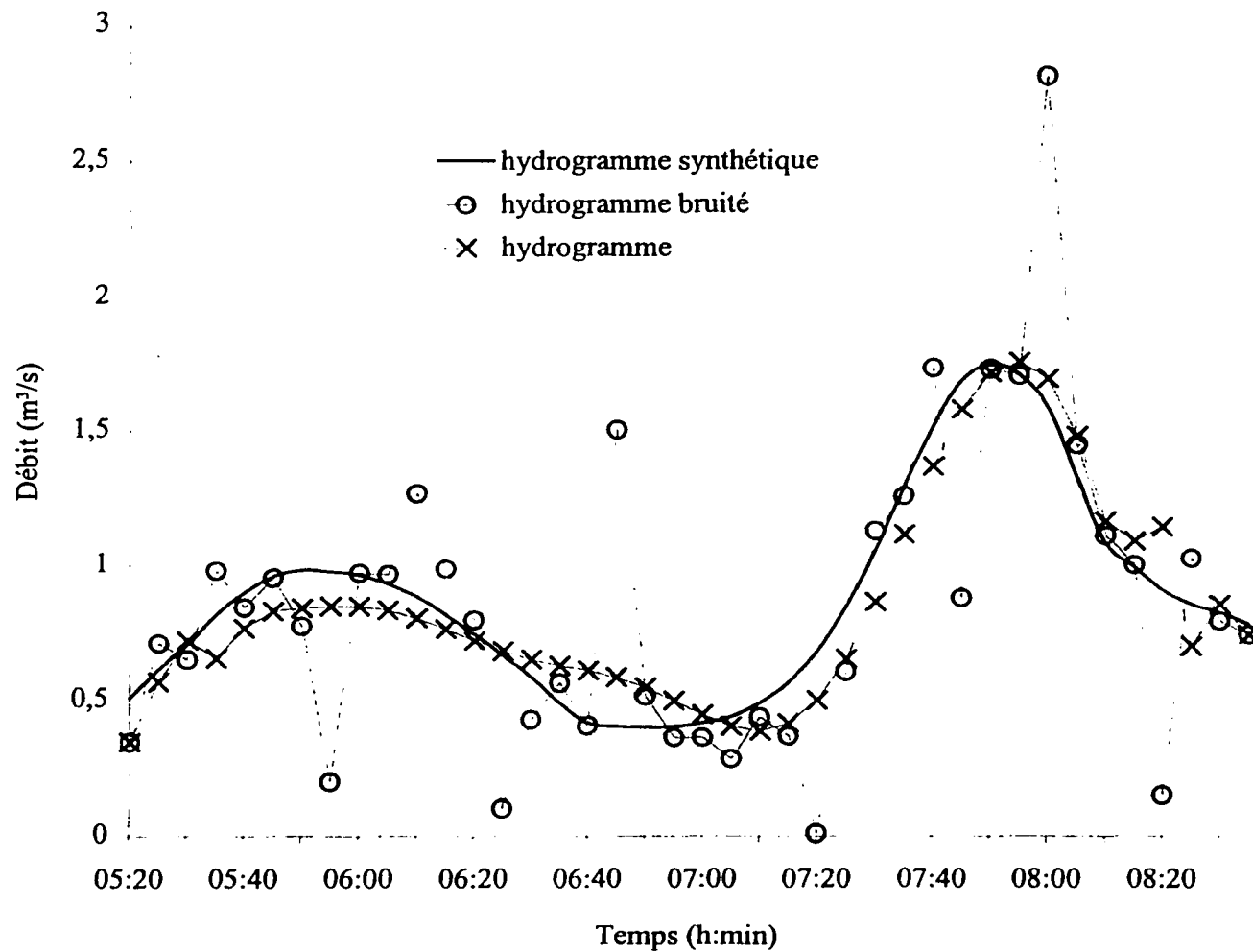


Figure 3.12 Filtrage de l'hydrogramme bruité- filtre à régression linéaire 7 puis fraser et fréquentiel 0.2

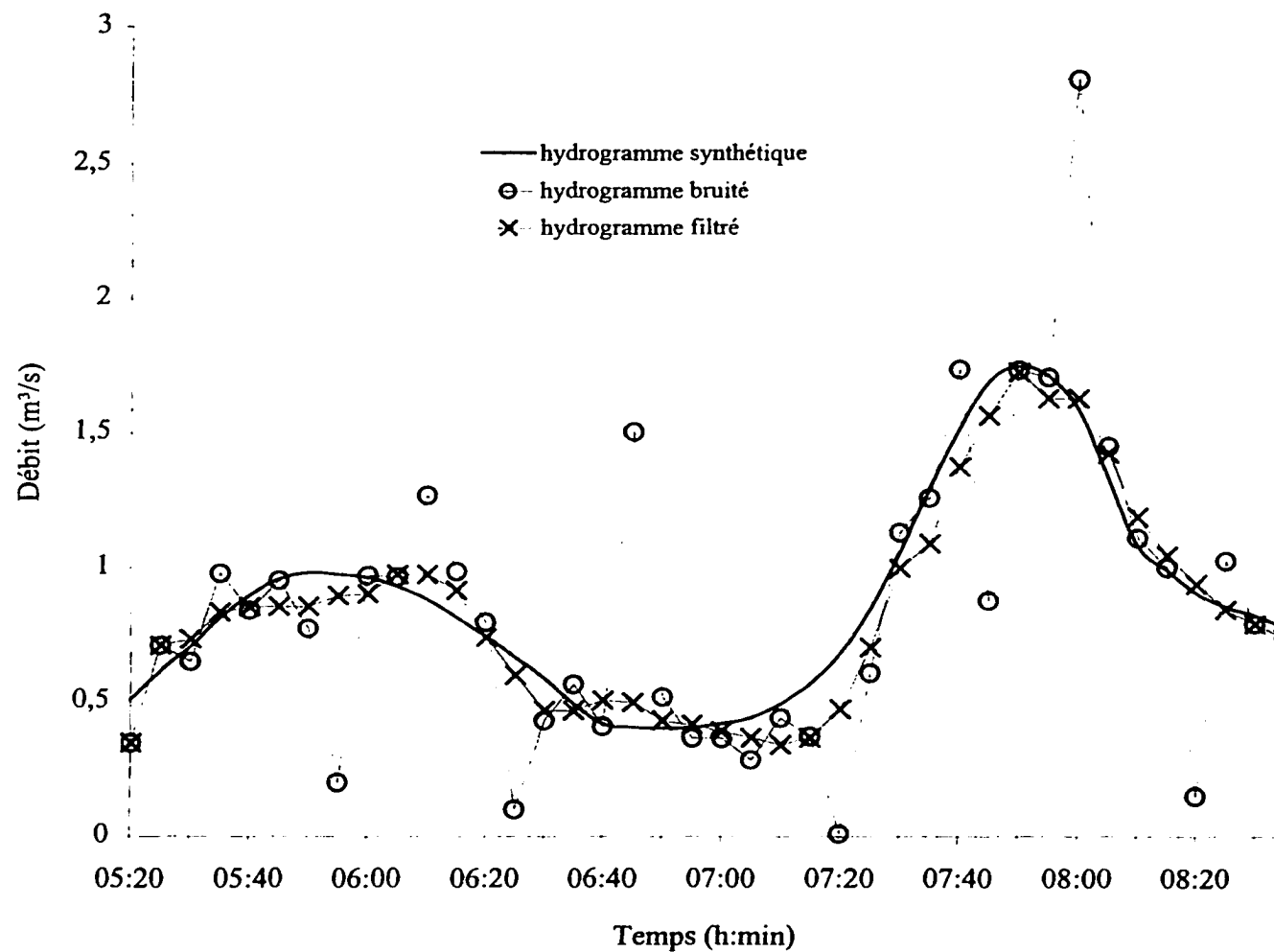


Figure 3.13 Filtrage de l'hydrogramme bruité- filtre winsoriseur de paramètres 5,2

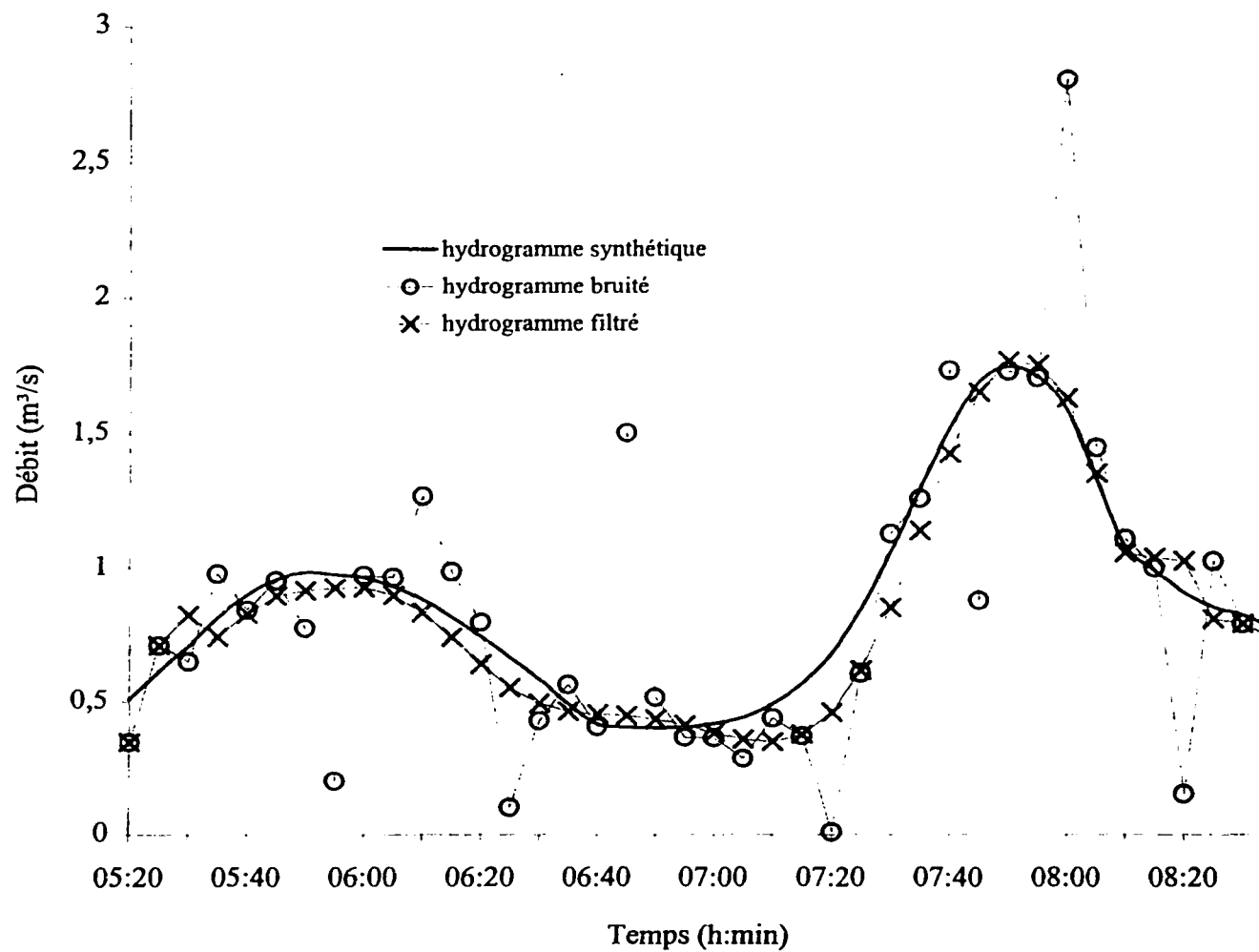


Figure 3.14 Filtrage de l'hydrogramme bruité- filtre winsoriseur 5,2 puis fraser et fréquentiel 0.2

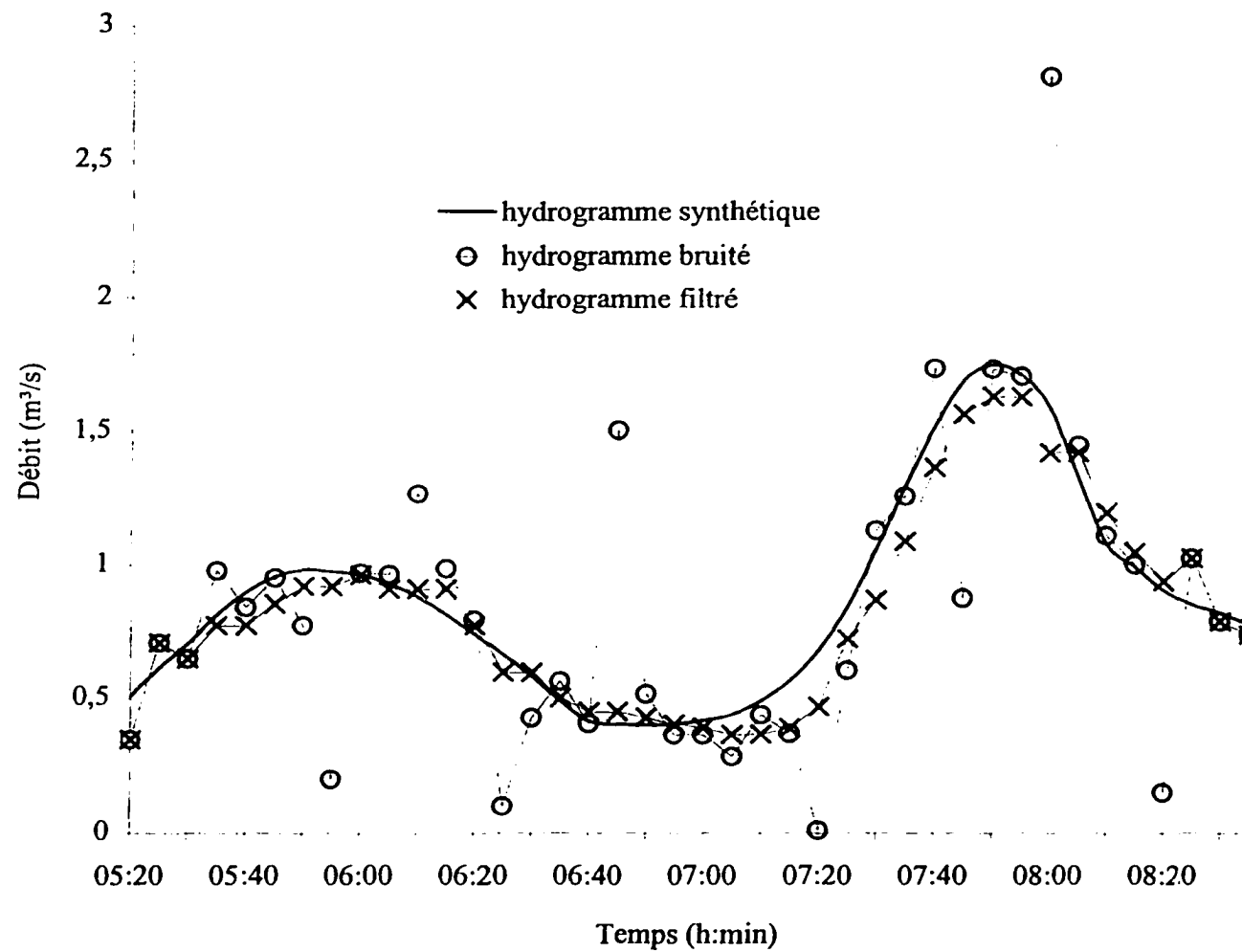


Figure 3.15 Filtrage de l'hydrogramme bruité- filtre winsoriseur de paramètres 7,4

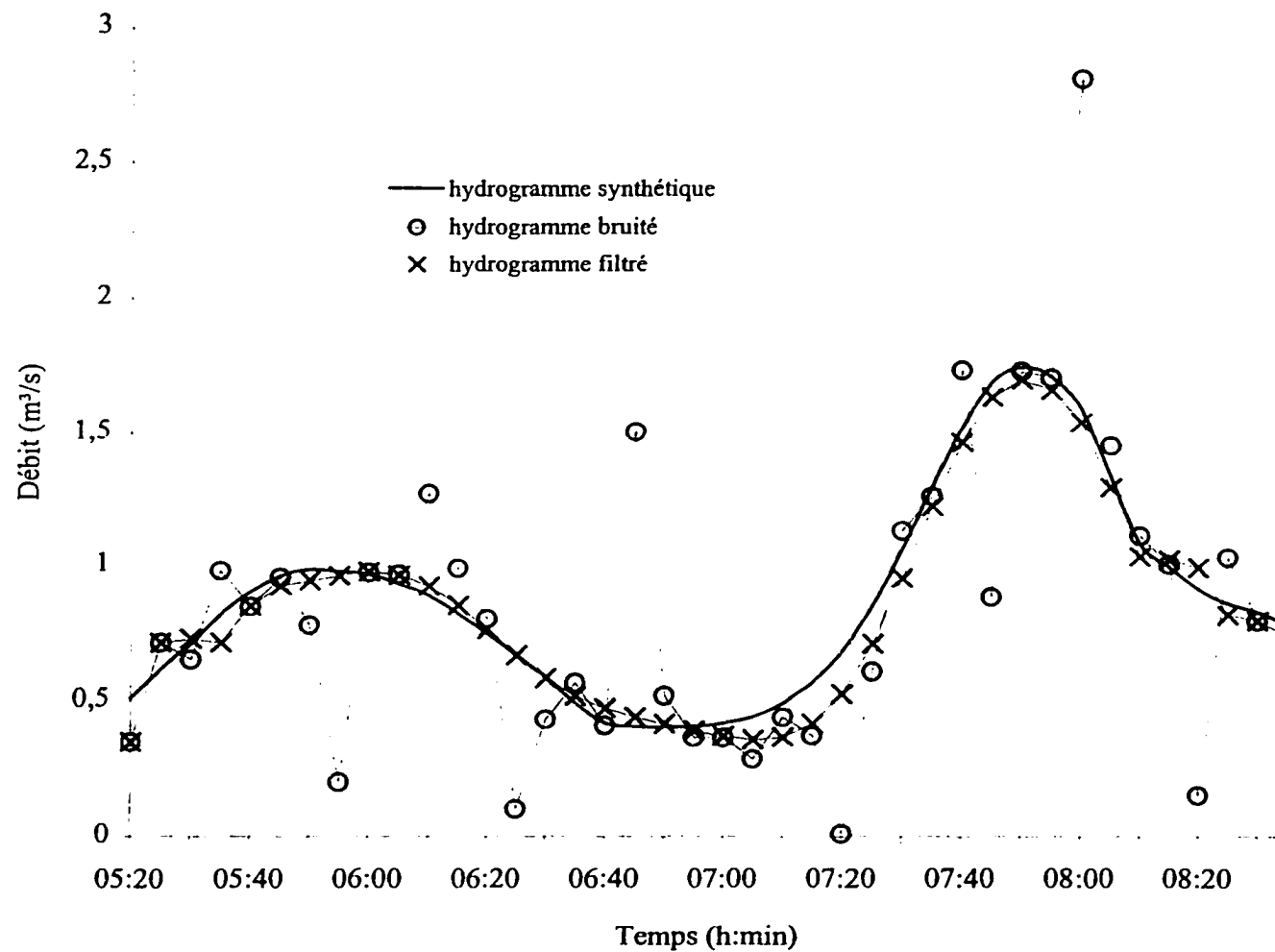


Figure 3.16 Filtrage de l'hydrogramme bruité- filtre winsoriseur 7,4 puis fraser et fréquentiel 0.2

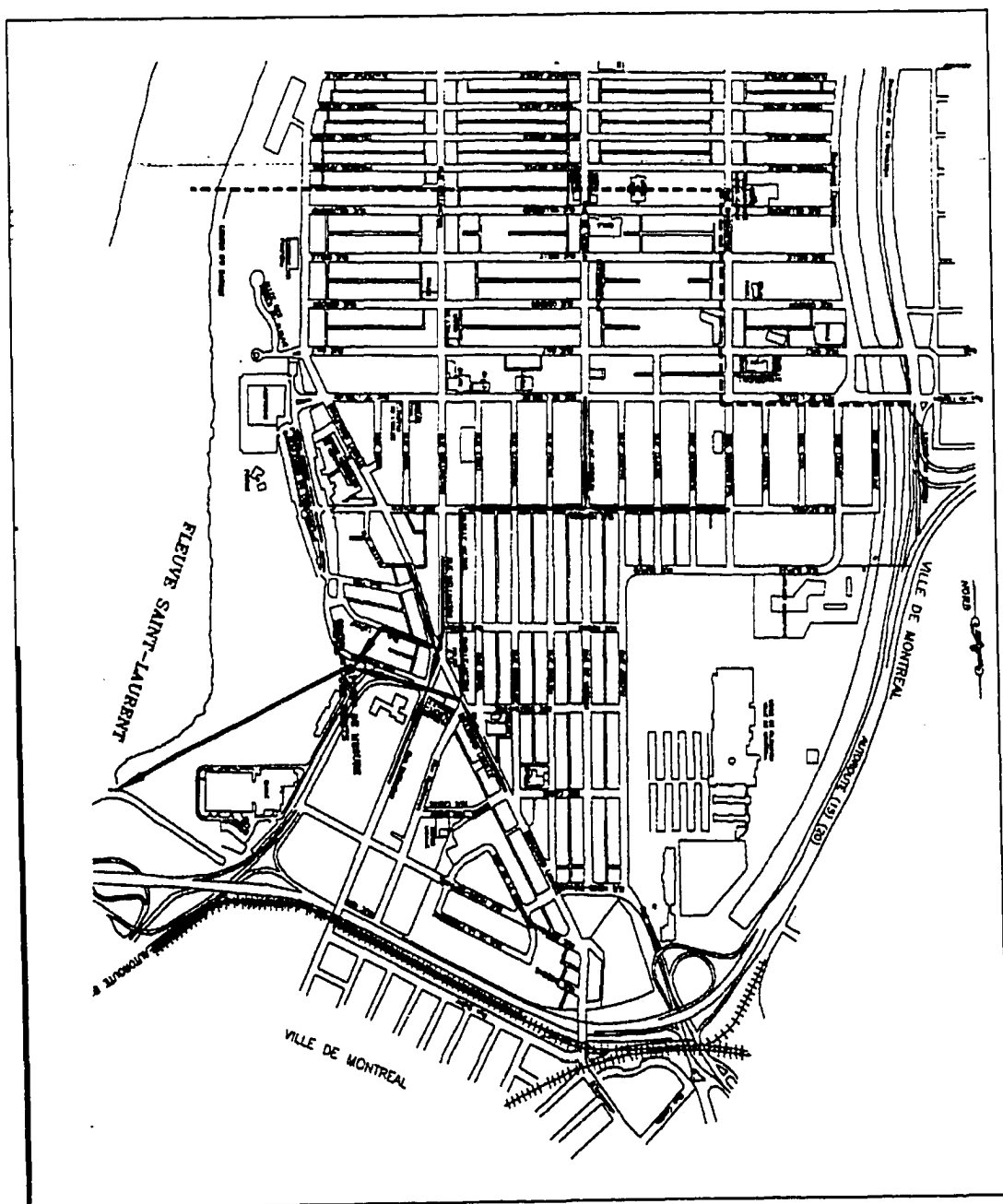


Figure 3.17 Schema du secteur I de la Ville de Verdun (Québec)

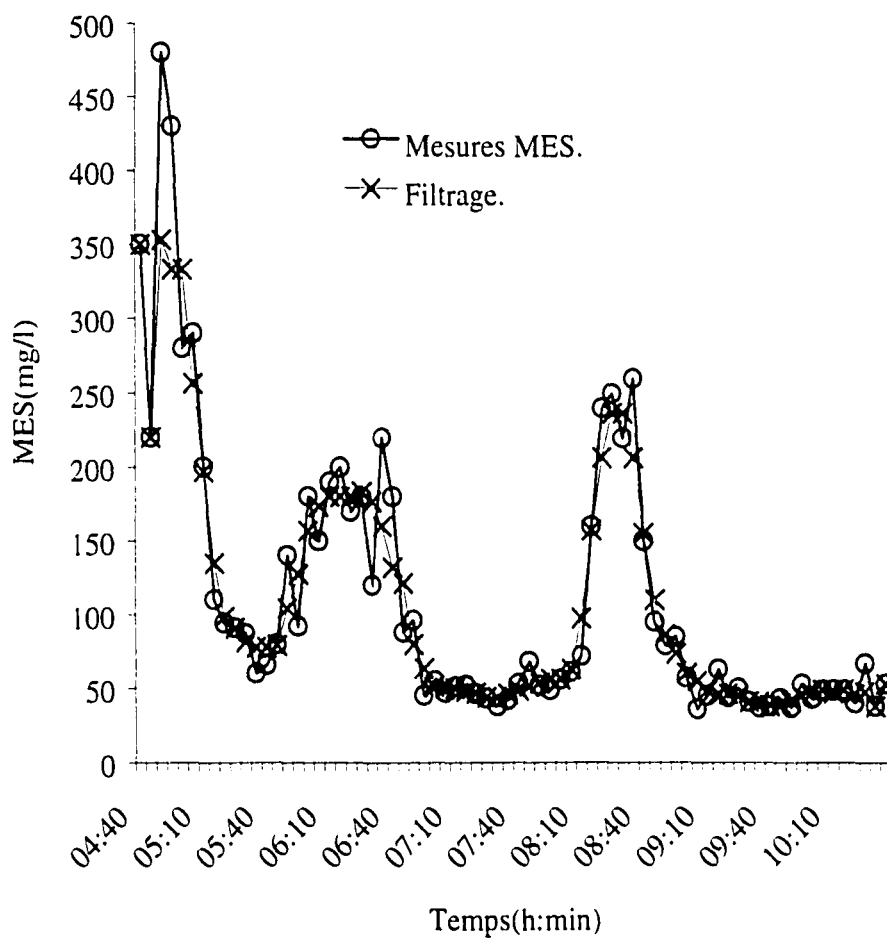


Figure 3.18 Filtrage de la série de mesures MES de Réhaume le 30/09/1999

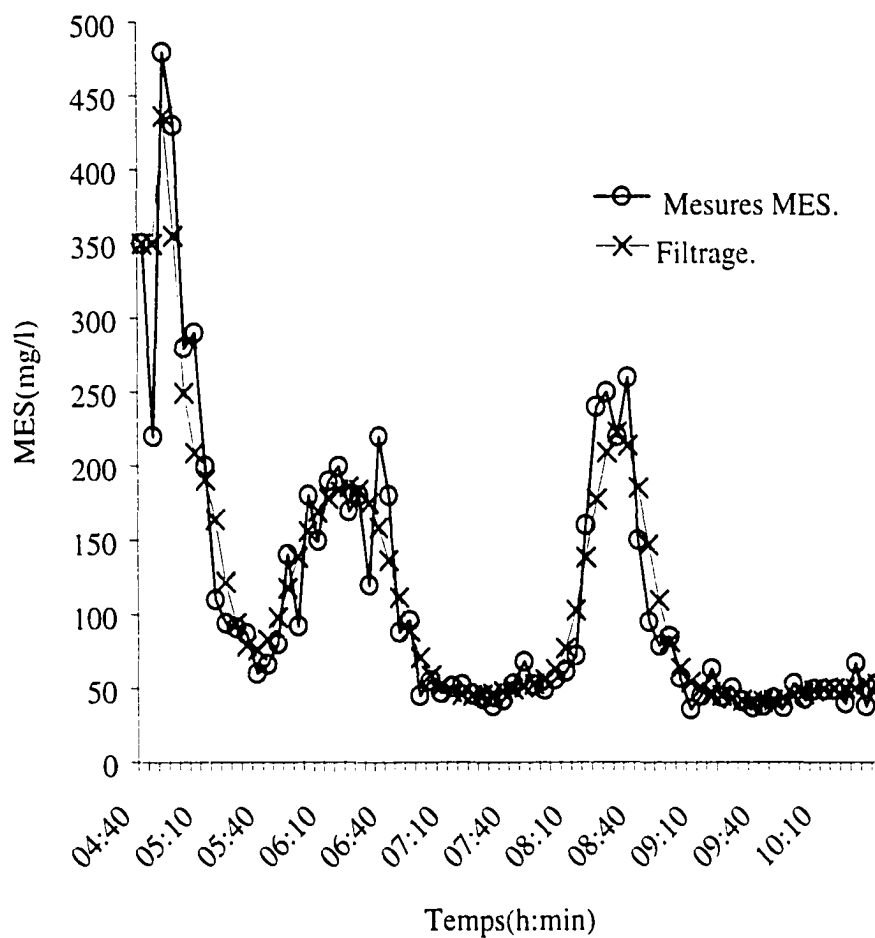


Figure 3.19 Filtrage de la série de mesures MES de Réhaume le 30/09/1999



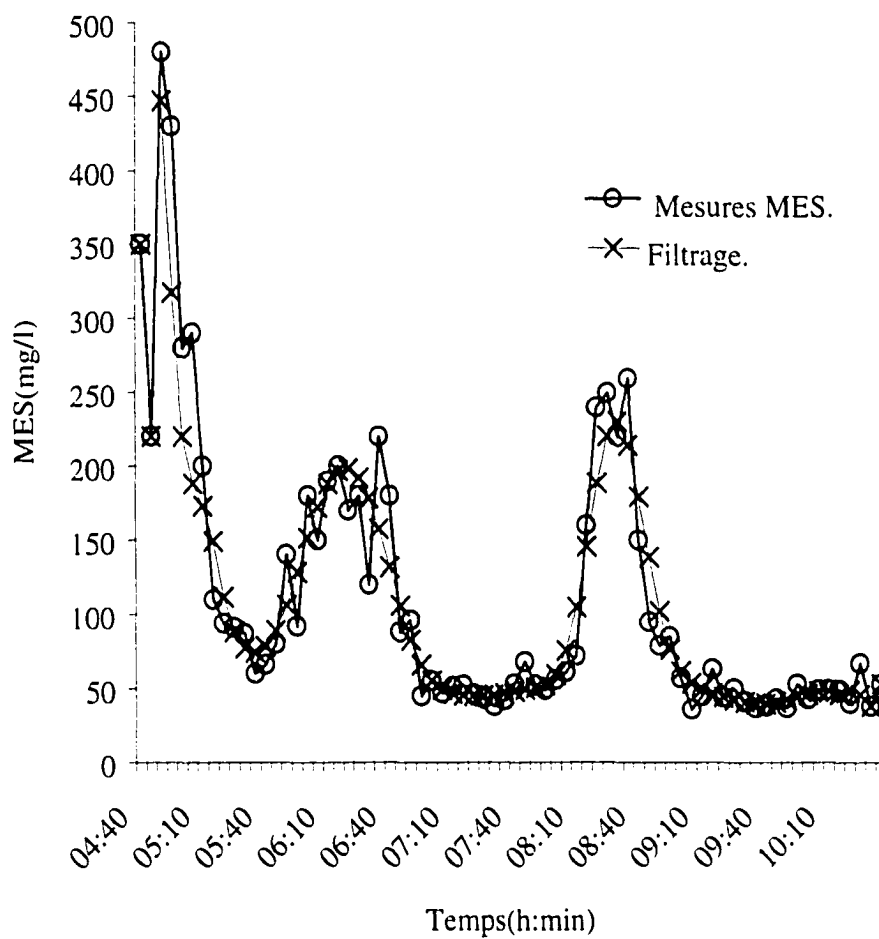


Figure 20: Filtrage de la série de mesures MES de Réhaume le 30/09/1999

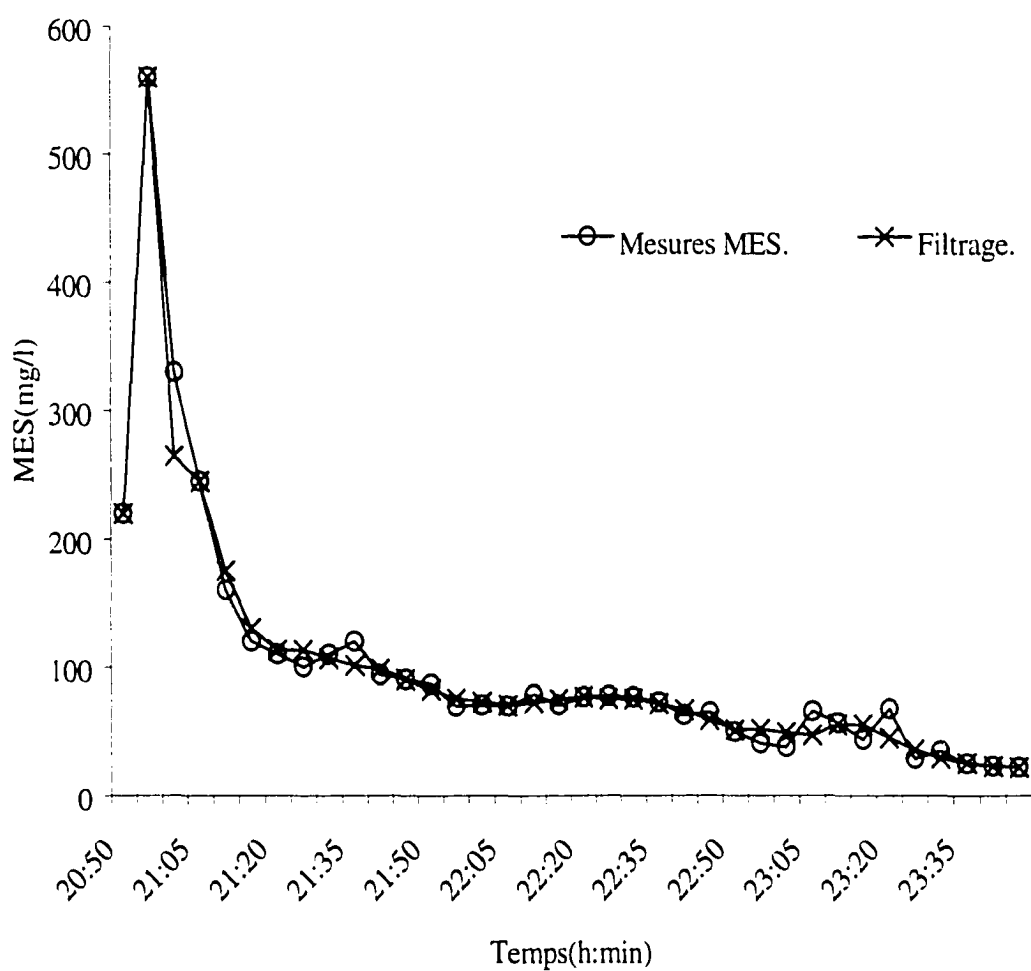


Figure 3.21 Filtrage 1 de la série de mesures MES de Rhéaume le 13/10/1999

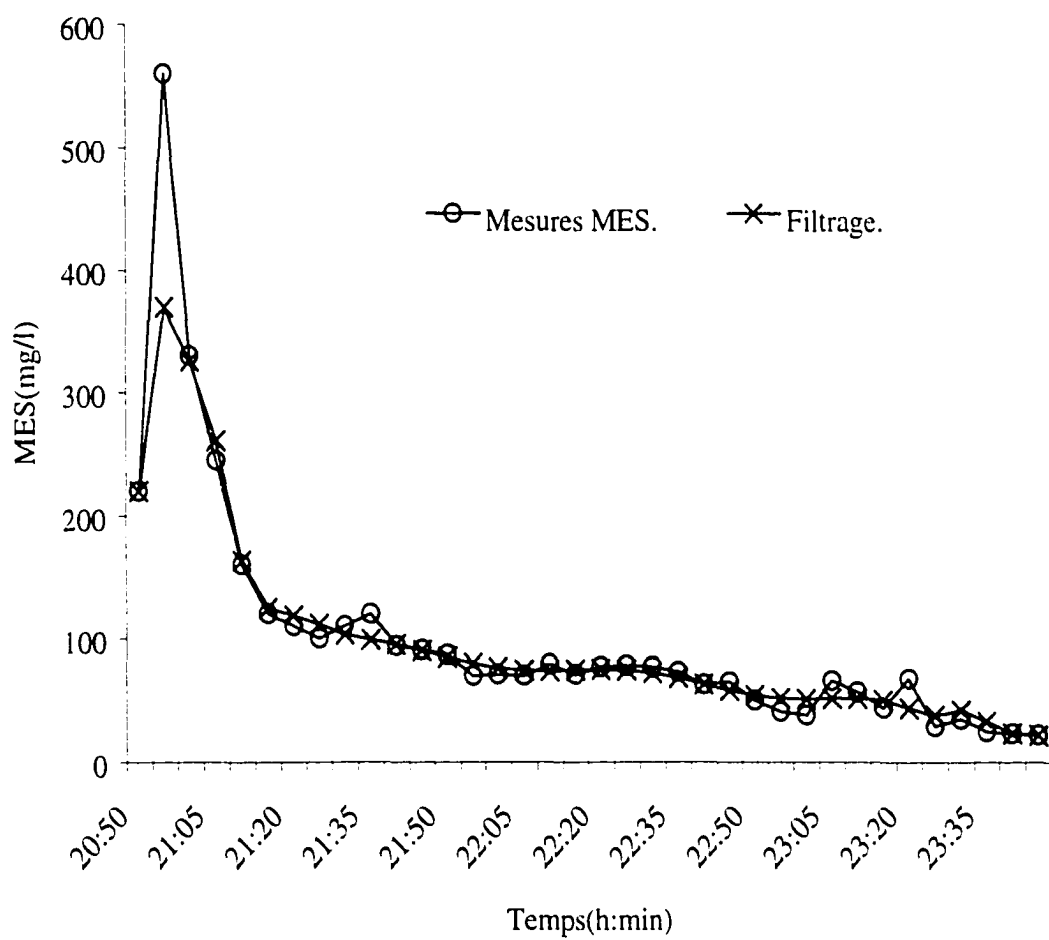


Figure 3.22 Filtrage 2 de la série de mesures MES de Rhéaume le 13/10/1999

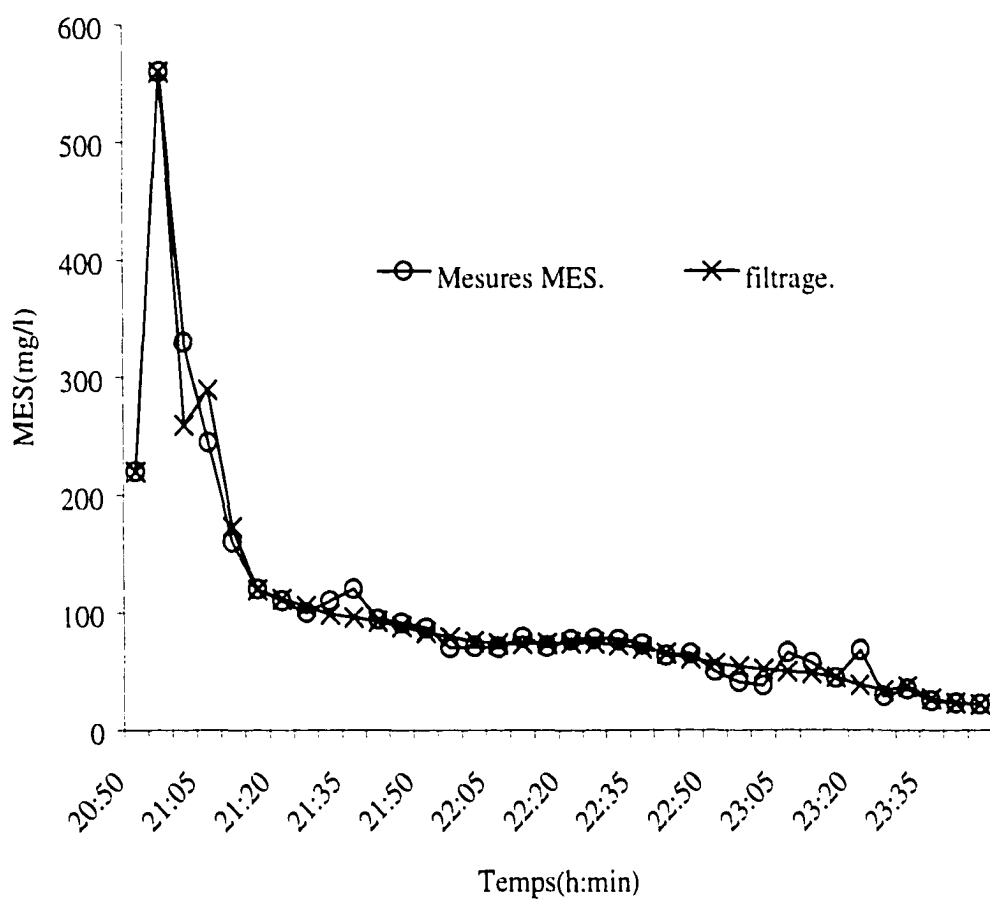


Figure 3.23 Filtrage 3 de la série de mesures MES de Rhéaume le 13/10/1999

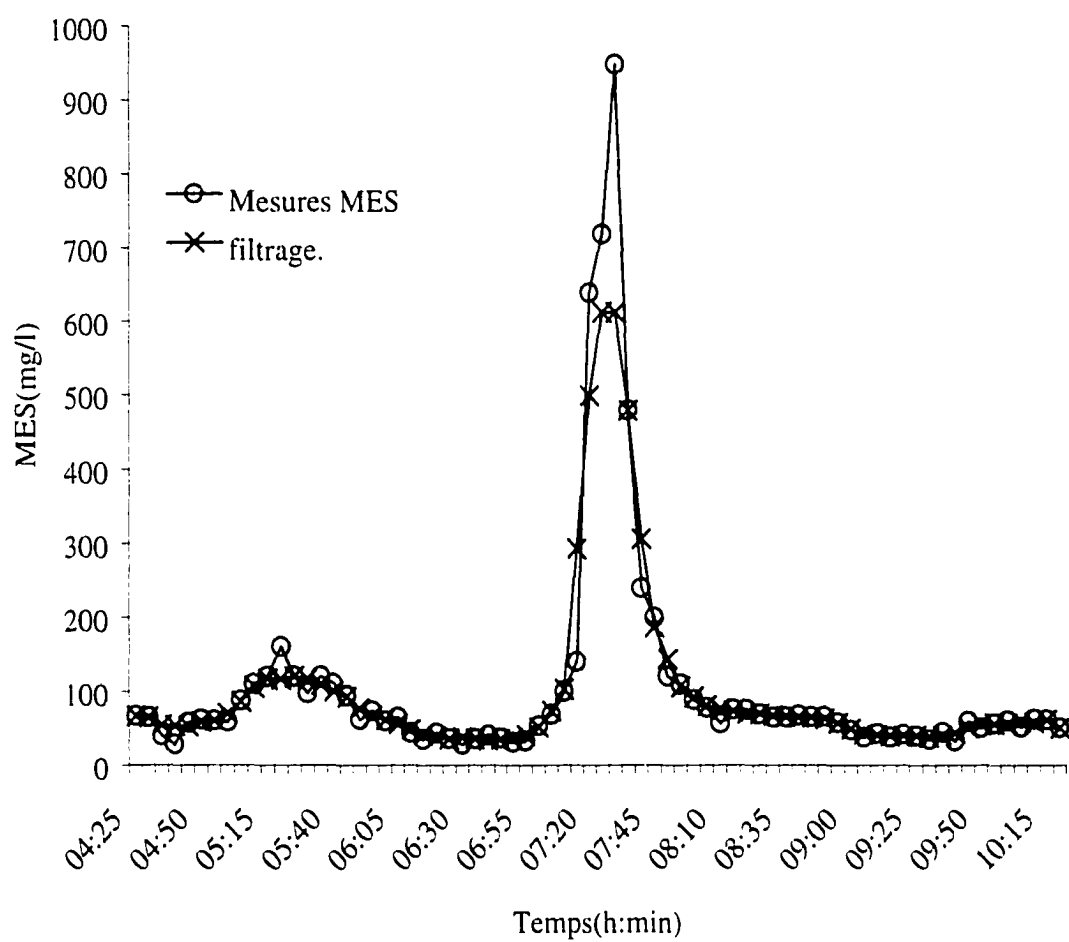


Figure 3.24 Filtrage I des mesures MES à Wellington le 30/09/1999

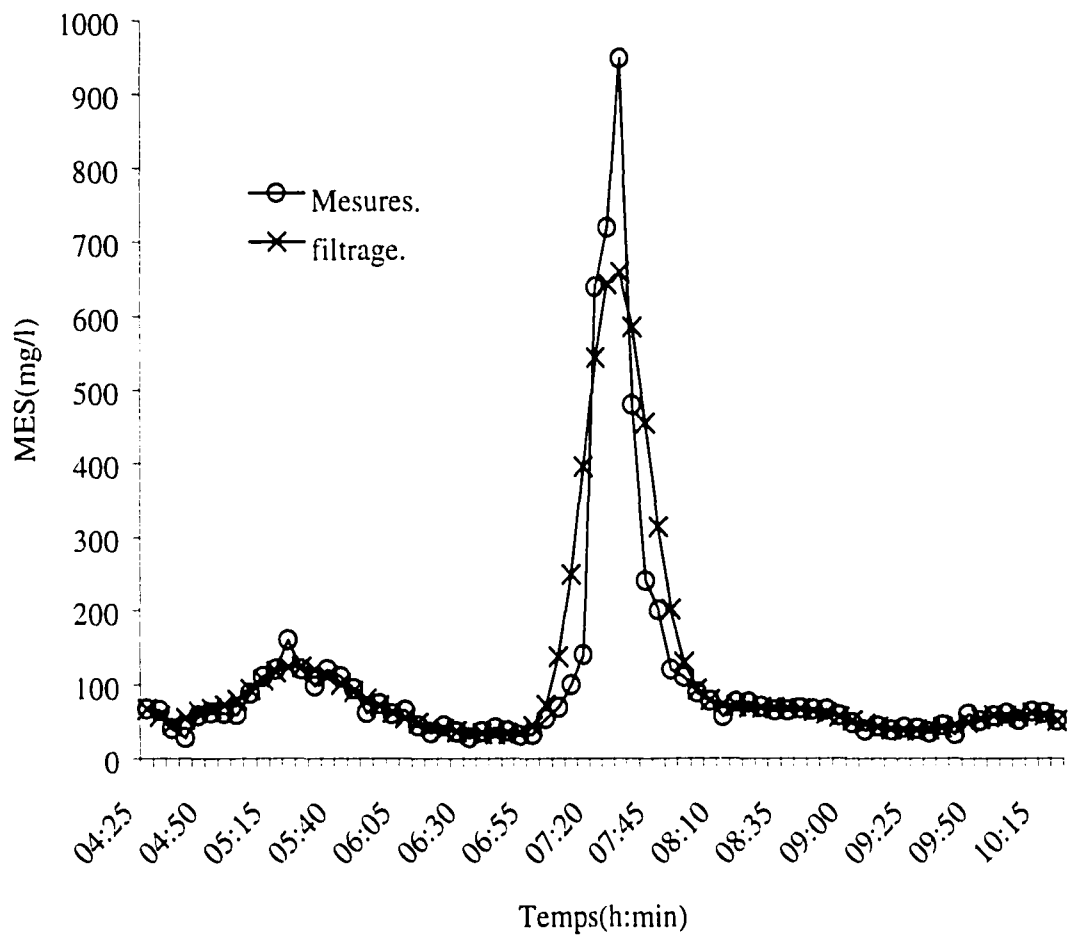


Figure 2 : filtrage 2 des mesures MES à Willington le 30/09/1999

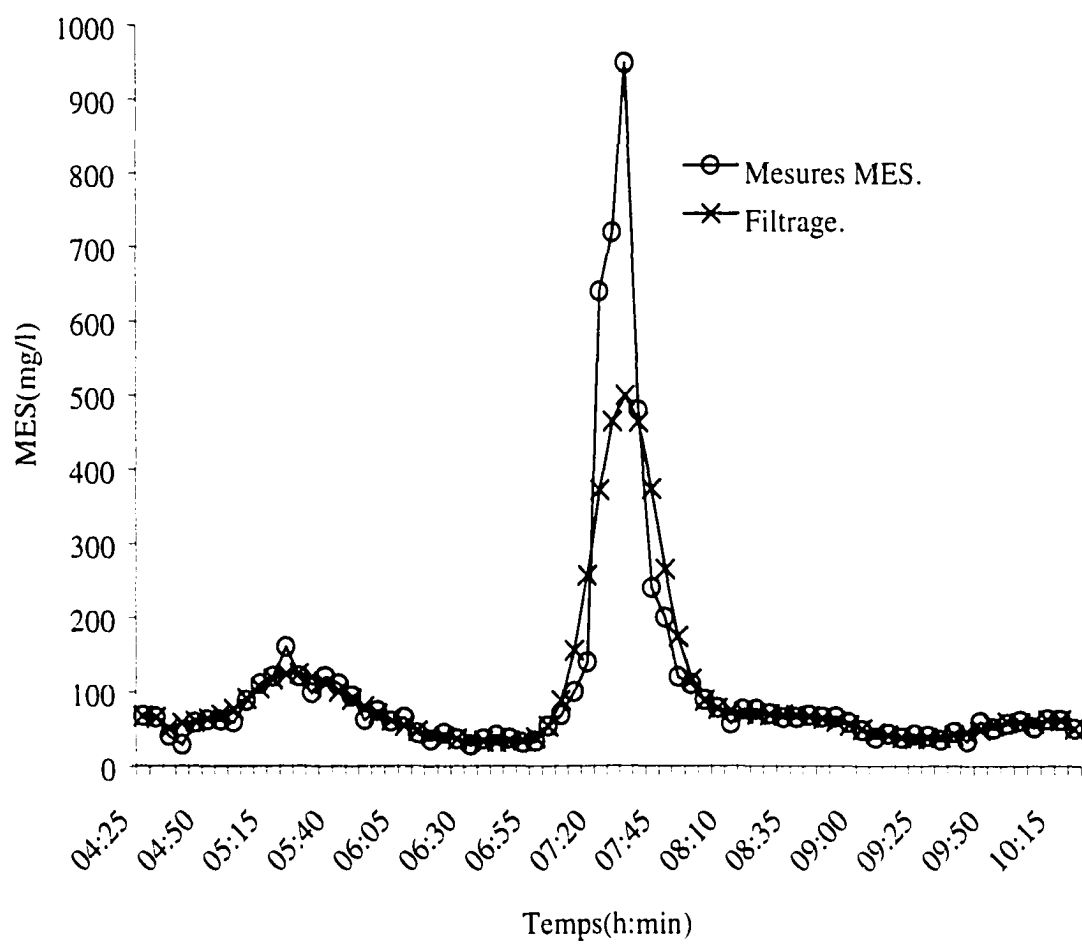


Figure 3.26 Filtrage 3 des mesures MES à Wellington le 30/09/1999

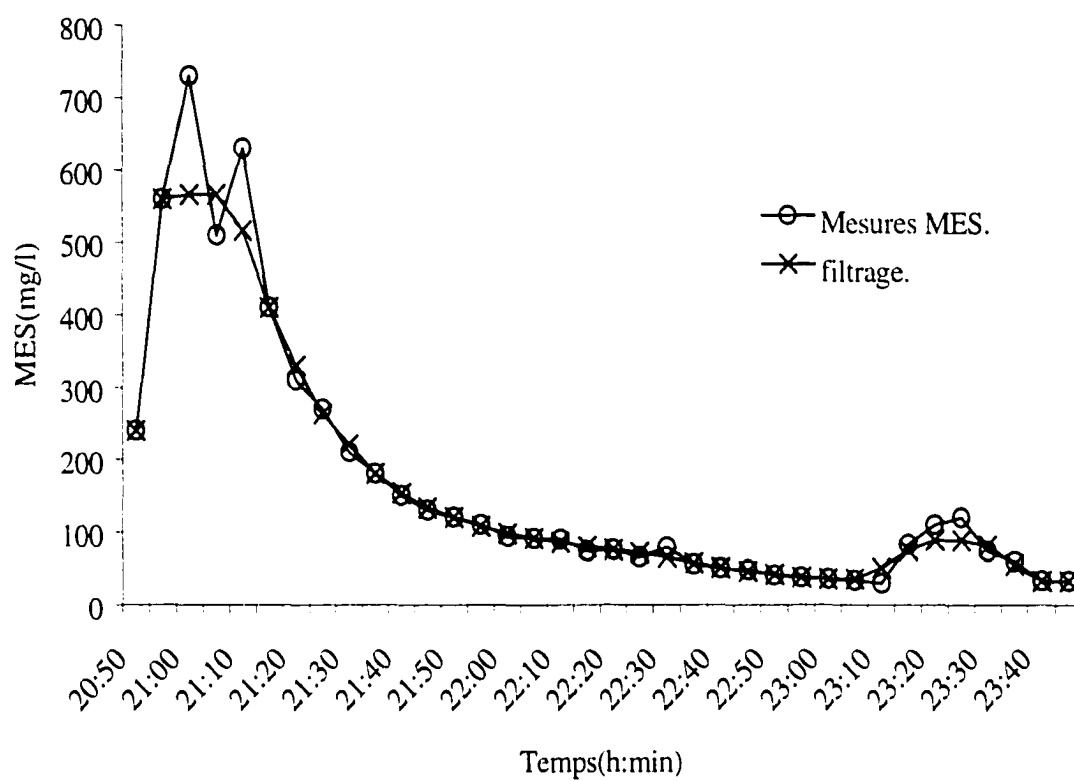


Figure 3.27 Filtrage 1 des mesures MES à Wellington le 13/10/1999.



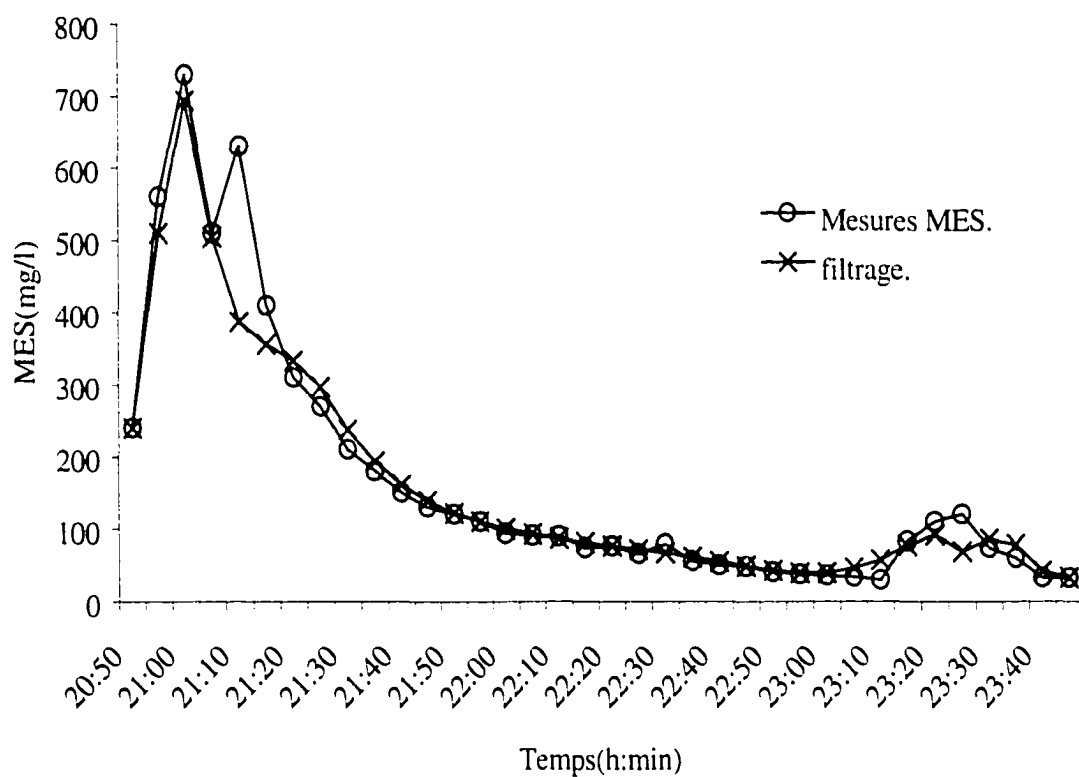


Figure 3.28 Filtrage 2 des mesures MES à Wellington le 13/10/1999.

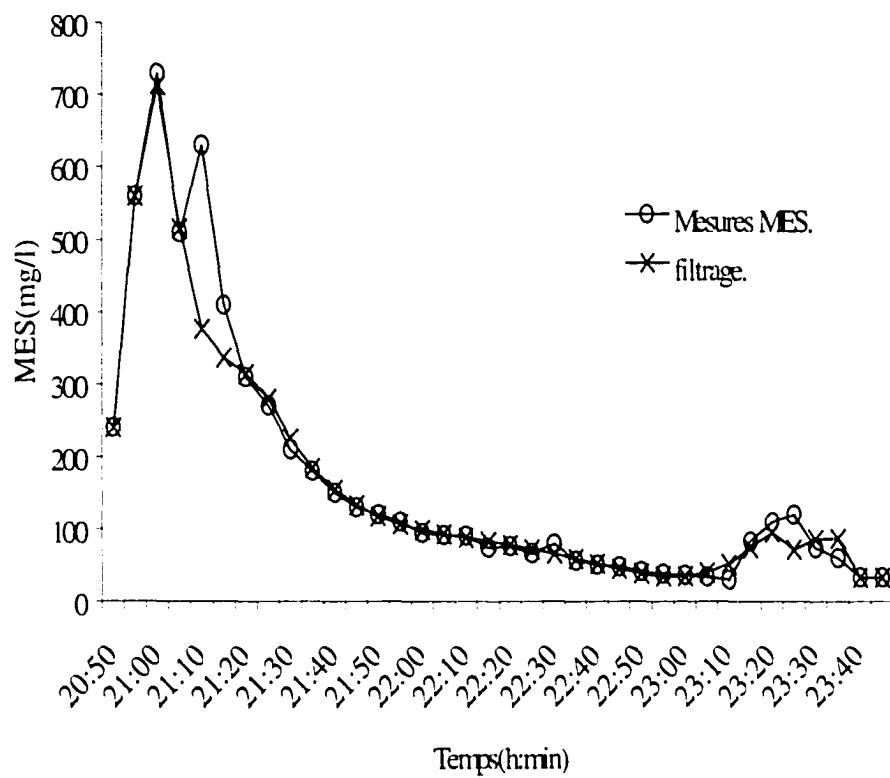


Figure 3.29 Filtrage 3 des mesures MES à Wellington le 13/10/1999.

## **CHAPITRE 4**

### **ESTIMATION DES DONNÉES MANQUANTES DE DÉBIT ET DE CONCENTRATION DE MATIÈRE EN SUSPENSION**

#### **4.1 Introduction**

De longues séries chronologiques de débit et de concentration de polluants sont nécessaires pour procéder à des études reliées à l'entretien, la réhabilitation, la conception et la gestion en temps réel des systèmes d'assainissement. Or en pratique, on ne dispose souvent que de séries de données partielles, manifestement insuffisantes pour une analyse hydraulique et environnementale complète. Dans le cas où des mesures existent, il est fréquent de constater des données erronées et parfois manquantes. Le filtrage et la validation des données ont fait l'objet de travaux antérieurs (Tahri et al, 2001), (Berrada et al., 1996) et (Bennis et Kang, 2000).

La littérature abonde de modèles d'estimation des données hydrométriques manquantes, alors que l'estimation des données de la qualité des eaux de ruissellement est beaucoup moins documentée. Ceci s'explique par le fait que des appareils de mesure, in situ, des concentrations de polluants ne sont apparues sur le marché que récemment.

Le modèle le plus utilisé est présenté sous forme de régression multiple où les données à reconstituer sont exprimées en fonction des données reconstituantes (Haan, 1977). Les paramètres du modèle régressif sont estimés à l'aide de la méthode classique des moindres carrés (Young, 1974). D'une façon générale, cette dernière technique donne de bons résultats. Cependant, elle possède trois inconvénients: (i) elle ne tient pas compte de la variation des paramètres de régression en fonction du temps; (ii) les variables dépendantes sont souvent corrélées entre elles, ce qui viole une des conditions de son application; (iii) les erreurs de mesure obéissent rarement à l'hypothèse de bruit gaussien utilisée dans la solution par la méthode standard des moindres carrés. Ce dernier point est la source de biais dans l'estimation des paramètres du modèle de régression utilisé (O'Connel, 1980).

D'autres algorithmes plus sophistiqués, comme les variables instrumentales et le maximum de vraisemblance, appliqués à des séries chronologiques entachées artificiellement d'un bruit aléatoire de moyenne nulle et de variance constante, produisent des solutions non biaisées meilleures que celles obtenues par la méthode standard des moindres carrés (Young, 1970). Malheureusement, ces algorithmes n'apportent plus d'amélioration significative lorsqu'il s'agit d'un signal expérimental entaché d'un bruit réel de caractéristiques statistiques inconnues qui ne répondent pas aux hypothèses idéales imposées (Davies, 1983).

L'analyse en composantes principales vient remédier à la dépendance mutuelle des variables indépendantes (McCuen et Snyder, 1986). Par une rotation des axes, on trouve de nouvelles variables orthogonales réellement indépendantes, appelées composantes principales. Celles-ci s'expriment sous forme de combinaisons linéaires des variables initiales. Les coefficients de pondération sont les éléments des vecteurs propres de la matrice de corrélation des débits.

La régression pseudo-orthogonale (Kachroo et Liang, 1992) peut aussi être utilisée pour remédier au problème de multicollinéarité des entrées du système.

Malheureusement, cette technique fait appel à un facteur arbitraire de points qu'il faut déterminer par essai-erreur, alors que notre objectif est de développer une méthodologie automatisée et générale d'application. De plus, le facteur de poids utilisé dans la régression pseudo-orthogonale introduit un biais dans la solution, laquelle n'est plus optimale au sens des moindres carrés. Ce biais, qui évolue dans le même sens que le facteur de poids, peut être modéré dans certaines situations (Bruen et Dooge, 1984), mais peut devenir très important pour d'autres applications (Kachroo et al., 1992).

Pour remédier au problème de stationnarité du modèle, nous utilisons l'algorithme récursif de moindres carrés, (Schilling et Martens, 1986). Cette technique offre les avantages d'un calcul itératif rapide, avec un espace mémoire très restreint dans l'ordinateur, et elle s'applique à des systèmes non stationnaires à entrées-sorties multiples.

Toutes les méthodes présentées jusqu'ici sont médiales ou à liaison spatiale, dans le sens où elles tiennent compte de l'information recueillie à l'extérieur de la station où il y a des données manquantes. Dans le cas où cette information ne serait pas disponible, on propose dans le présent travail des méthodes immédiates qui n'utilisent que les informations à la station reconstituée. Il s'agit, dans ce cas, de modèles univariés. Les modèles ARIMA (Box et Jenkins, 1976) sont appropriés pour représenter l'évolution des débits et des concentrations et reconstituer les données manquantes.

## 4.2 Méthodes d'estimation des données manquantes

### 4.2.1 Introduction

Les méthodes de reconstitution des données sont à liaison spatiale si elles tiennent compte de l'information provenant de l'extérieur de la station où il existe des données manquantes. Souvent, cette information n'est pas disponible. On propose donc, de se limiter aux informations recueillies dans la station à reconstituer. Il s'agit des méthodes de reconstitution univariées.

Les modèles de type Box et Jenkins ARMA qui sont le plus souvent utilisés pour la prévision, seront utilisés dans ce mémoire pour l'estimation des données manquantes

### 4.2.2 Méthode univariée d'estimation des données manquantes

#### 4.2.2.1 Modèles de prévision

Parmi les modèles ARMA (m,p,n) le modèle autorégressif est certainement le plus populaire pour la prévision. Ainsi, pour la prévision des débits d'écoulement ou la concentration d'un polluant le modèle AR(m) s'écrit :

$$\begin{cases} Q_t = \sum_{j=1}^m b_j \times Q_{t-j} + W_t \\ C_t = \sum_{j=1}^n d_j \times C_{t-j} + U_t \end{cases} \quad (4.1)$$

où :  $b_j$  et  $d_j$  sont les coefficients du modèle autorégressif  $n$  et  $m$  sont les ordres des modèles.

$U_t$  et  $W_t$  sont des bruits qui tiennent compte des erreurs de mesure et modélisation.

En pratique, on utilise les mesures prises jusqu'à l'instant «  $t$  » pour identifier les paramètres du modèle  $b_j$  et  $d_j$ . les prévisions pour les instants  $t+1$ ,  $t+2$ ,...sont obtenues

par extrapolation du modèle (1) en supposant que  $W_{t+1}$ ,  $W_{t+2}$ , et  $U_{t+1}$ ,  $U_{t+2}, \dots$  sont nuls.

La problématique d'estimation des données manquantes est différente du problème de prévision. Lorsqu'il s'agit d'estimer des données manquantes historiques on dispose de données mesurées aussi bien avant qu'après la période sur laquelle on estime les données. Il est ainsi possible d'extrapoler la série chronologique mesurée de part et d'autre des données mesurées pour estimer les données manquantes (figure 4.1). Le modèle autorégressif dans le sens aller «A » s'écrit :

$$\begin{cases} Q_t^A = \sum_{j=1}^m b_j^A Q_{t-j}^A + W_t^A \\ C_t^A = \sum_{j=1}^n d_j^A C_{t-j}^A + U_t^A \end{cases} \quad (4.2)$$

Le modèle autorégressif dans le sens retour « R » s'écrit :

$$\begin{cases} Q_t^R = \sum_{j=1}^m b_j^R Q_{t+j}^R + W_t^R \\ C_t^R = \sum_{j=1}^n d_j^R C_{t+j}^R + U_t^R \end{cases} \quad (4.3)$$

$C_t^A$  : représente l'estimation de la concentration en polluant « Aller »;

$C_t^R$  : représente l'estimation de la concentration en polluant «Retour »;

$Q_t^A$  : représente l'estimation du débit «Aller »;

$Q_t^R$  : représente l'estimation du débit «Retour »;

$U_j^A$ ,  $W_t^A$  : représentent les bruits du modèle «Aller »;

$U_j^R$ ,  $W_t^R$  : représentent le bruit du modèle «Retour ».

L'ordre optimal du modèle AR peut être déterminé, pour chaque série, par l'usage du critère d'Akaike (Akaike, 1974). Comme nous sommes intéressés par un programme

automatisé d'estimation des données hydrométriques et des concentrations en polluants, nous avons arbitrairement fixé l'ordre du modèle AR à (2). Ceci, est justifié par le besoin d'utiliser des modèles parcimonieux (Abraham and Ledolter, 1983). La présentation du modèle AR(2) utilisé ici a été adaptée à la formulation générale de l'algorithme du filtre de Kalman (Fitch and McBean, 1991). Le modèle AR(2) se présente sous la forme suivante :

$$\begin{cases} Q_t = b_1 \times Q_{t-1} + b_2 \times Q_{t-2} + W_{t-1} \\ C_t = d_1 \times C_{t-1} + d_2 \times C_{t-2} + U_{t-1} \end{cases} \quad (4.4)$$

Où  $\begin{cases} Q_t \\ C_t \end{cases}$  représentent respectivement le débit et la concentration en polluant à l'instant  $t$

et  $W_{t-1}, U_{t-1}$  sont les bruits du système. Ceci, suppose que nous avons observé  $\begin{cases} Q_t \\ C_t \end{cases}$  avec

un bruit de mesure  $(V_t, Y_t)$  :

$$\begin{cases} Z_t = Q_t + V_t \\ X_t = C_t + Y_t \end{cases} \quad (4.5)$$

Soit :

$$\begin{aligned} \begin{cases} X_t = [Q_t & Q_{t-1}]^T \\ Y_t = [C_t & C_{t-1}]^T \end{cases} \text{ et } \phi = \begin{bmatrix} b_1 & b_2 \\ 1 & 0 \end{bmatrix} \quad \psi = \begin{bmatrix} d_1 & d_2 \\ 1 & 0 \end{bmatrix} \\ W_{t-1} = [W_{t-1} \quad 0] \quad H_t = [1 \quad 0], \quad U_{t-1} = [U_{t-1} \quad 0] \quad H_t = [1 \quad 0] \end{aligned}$$

ainsi :

$$\begin{cases} \begin{bmatrix} Q_t \\ Q_{t-1} \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} Q_{t-1} \\ Q_{t-2} \end{bmatrix} + \begin{bmatrix} W_{t-1} \\ 0 \end{bmatrix} \\ \begin{bmatrix} C_t \\ C_{t-1} \end{bmatrix} = \begin{bmatrix} d_1 & d_2 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} C_{t-1} \\ C_{t-2} \end{bmatrix} + \begin{bmatrix} U_{t-1} \\ 0 \end{bmatrix} \end{cases} \quad (4.6)$$



et

$$\begin{cases} Z_t = [1 & 0] \times \begin{bmatrix} Q_t \\ Q_{t-1} \end{bmatrix} + V_t \\ X_t = [1 & 0] \times \begin{bmatrix} C_t \\ C_{t-1} \end{bmatrix} + Y_t \end{cases} \quad (4.7)$$

L'équation (4.6) et (4.7) peuvent être écrites comme suit :

$$\begin{cases} X_t = \phi_{t,t-1} \times X_{t-1} + W_{t-1} \\ Y_t = \psi_{t,t-1} \times Y_{t-1} + U_{t-1} \end{cases} \quad (4.8)$$

$$\begin{cases} Z_t = H \times X_t + V_t \\ X_t = H \times X_t + Y_t \end{cases} \quad (4.9)$$

L'équation (4.8) représente le modèle AR dans les sens croissants du temps. L'équation pour le modèle AR dans le sens rétrograde du temps est obtenue par inversion de t et t-1 dans la même équation. À chaque temps t, l'algorithme FK est utilisé : la première fois pour filtrer les mesures la deuxième fois pour identifier les paramètres du modèle qu'on suppose variables. Ce travail est fait pour les deux modèles AR : aller et retour.

#### 4.2.2.2 Technique de lissage de Fraser

Avant de présenter la technique d'estimation des données manquantes pour le cas univarié, on rappelle la technique de lissage de Fraser (Radix, 1984). Pour lisser les données, l'algorithme de Fraser fait une estimation optimale de la valeur au temps t des N+1 mesures prises aux temps  $t_0, t_1, t_2, \dots, t_N$ .

La technique de filtrage de Fraser consiste en trois étapes (Radix, 1984):

- ◆ Appliquer un modèle de prévision, en l'occurrence AR, de  $t_0$  à  $t$  et utiliser les conditions initiales  $P_{t_0}$  et, où  $P_{t_0}$  est la matrice de covariance de l'estimation de l'erreur de  $X_{t_0}$  dans le but d'obtenir  $\hat{P}_t^A$  et  $\hat{X}_t^A$ .
- ◆ Appliquer l'algorithme FK combiné avec le modèle AR dans le sens rétrograde du temps en allant de  $t_N$  à  $t$  en utilisant les conditions initiales  $P_{t_N}$  et  $X_{t_N}$  pour obtenir  $\hat{P}_t^R$  et  $\hat{X}_t^R$ .
- ◆ Calculer une combinaison optimale des deux estimations utilisant la relation suivante :

$$\hat{P}_t^{-1} = (\hat{P}_t^A + \hat{P}_t^R)^{-1} \quad (4.10)$$

$$\hat{X}_t = \hat{P}_t \times \left[ (\hat{P}_t^A)^{-1} \times \hat{X}_t^A + (\hat{P}_t^R)^{-1} \times \hat{X}_t^R \right] \quad (4.11)$$

L'estimation optimale est, désormais, une combinaison linéaire des estimations aller et retour.

#### 4.2.2.3 Technique d'interpolation

Nous proposons d'adopter la même méthodologie pour estimer les données manquantes en utilisant deux modèles AR. Le premier modèle AR qui fonctionne dans le sens normal du temps est exprimé comme suit :

$$X_t^A = \phi_{t/t-1} \times X_{t-1}^A + W_{t-1}^A. \quad (4.12)$$

L'exposant A désigne la direction aller dans le temps.

Le deuxième modèle AR qui fonctionne dans le sens rétrograde du temps est exprimé comme suit :

$$X_t^R = \phi_{t/t-1} \times X_{t-1}^R + W_{t-1}^R \quad (4.13)$$

L'exposant R désigne la direction retour dans le temps.

Chaque donnée manquante est estimée deux fois. Une première fois par extrapolation du modèle (4.12) dans le sens croissant du temps et une seconde fois par extrapolation du modèle (4.13) dans le sens rétrograde du temps (figure 4.1).

Afin de comprendre le fonctionnement des modèles, on considère une série de données mesurées à partir de l'instant «0 » à «t », une succession de données manquantes de «t+1 » à «t+M », et une série de données mesurées disponibles de «t+M+1 » à «t<sub>f</sub> ».

Le filtre de Kalman, utilisé selon la technique MISP «Mutually interactive state parameters estimation », permet après calibration sur les données de l'instant 0 à t, d'estimer les données manquantes par extrapolation du modèle autorégressif aux instants t+1, t+2,...,t+M à l'aide du modèle (4.12). D'une manière similaire, le modèle (4.13) permet, après calibration sur les données de l'instant «t<sub>f</sub> » à «t+M+1 », d'estimer les débits ou concentrations en MES manquants pour les mêmes périodes de temps en prenant comme condition initiale t<sub>f</sub> et en utilisant le modèle (4.13). Pour chaque instant «t » compris entre «t+1 » et «t+M+1 » on dispose donc de deux estimations différentes  $\begin{Bmatrix} \hat{Q}_t^A \\ \hat{C}_t^A \end{Bmatrix}$  et  $\begin{Bmatrix} \hat{Q}_t^R \\ \hat{C}_t^R \end{Bmatrix}$  de chaque donnée manquante. Pour utiliser ces deux estimations, il faudrait les pondérer par l'intermédiaire de leur erreurs d'estimation respectives  $e_t^A$  et  $e_t^R$  (Winkler et Makridakis, 1983) :

$$\hat{Q}_t = W_1 \times Q_t^A + W_2 \times Q_t^R \quad (4.14)$$

si l'on désigne par  $\Sigma$  la matrice de covariance des erreurs d'estimation des méthodes « Aller » et « Retour », les poids  $W_i$  sont exprimés par la relation :

$$W_i = \sum_{j=1}^2 \alpha_{ij} / \sum_{h=1}^2 \sum_{j=1}^2 \alpha_{hj} \quad (4.15)$$

où les  $\alpha_{ij}$  sont des termes de la matrice  $\Sigma^{-1}$ , les termes de la matrice  $\Sigma$  étant de la forme:

$$\begin{cases} (\Sigma)_{ij} = \text{cov}[e_t^i, e_t^j] & \text{pour } i \neq j \\ (\Sigma)_{ii} = \text{var}[e_t^i] \end{cases} \quad \begin{cases} e_{1t}^i = \frac{Q_t - \hat{Q}_t^i}{Q_t} \\ e_{2t}^i = \frac{C_t - \hat{C}_t^i}{C_t} \end{cases} \quad (4.16)$$

Le filtre de Kalman ne fournit que l'erreur d'estimation de la première prévision. Les erreurs d'extrapolation d'avance  $l$  par un modèle ARMA peuvent être obtenues en mettant le modèle AR sous forme (Pankratz, 1983) :

$$\begin{cases} Q_t = \bar{Q} + \psi_0 \times a_t + \psi_1 \times a_{t-1} + \psi_2 \times a_{t-2} + \psi_3 \times a_{t-3} \dots \\ C_t = \bar{C} + \varphi_0 \times b_t + \varphi_1 \times b_{t-1} + \varphi_2 \times b_{t-2} + \varphi_3 \times b_{t-3} \dots \end{cases} \quad (4.17)$$

La variance de l'erreur d'extrapolation d'avance «  $l$  » s'obtient par :

$$\sigma^2[e_t(l)] = \sigma_a^2 \times (1 + \psi_1^2 + \psi_2^2 \dots + \psi_{l-1}^2) \quad (4.18)$$

où  $a_t$  est un bruit gaussien de moyenne nulle et de variance  $\sigma_a^2$ ,  $\psi_i$  sont des constantes reliées aux coefficients du modèle ARMA.

Les étapes à suivre pour estimer les données manquantes sont les suivantes :

- Un filtrage de Kalman aller, de  $t_1$  à  $t_i$ , permettant le calcul des coefficients autorégressifs représentés par le vecteur :  $\hat{\Theta}_{i/i}^A$  ;
- Une prévision aller, selon la méthode ARMA en utilisant les coefficients du vecteur  $\hat{\Theta}_{i/i}^A$ . On calcule ainsi, l'estimation aller,  $\hat{z}_j^A$  et la variance aller  $\hat{\sigma}_j^A$  de la  $j^{\text{ème}}$  donnée manquante;

- c. Un filtrage de Kalman retour, de  $t_N$  à  $t_{m+1}$ , permettant le calcul des coefficients autorégressifs représentés par le vecteur :  $\hat{\Theta}_{m+1/m+1}^R$  ;
- d. Une prévision retour, selon la méthode ARMA en utilisant les coefficients du vecteur  $\hat{\Theta}_{m+1/m+1}^R$ . On calcule ainsi, l'estimation retour,  $\hat{z}_j^R$  et la variance retour  $\hat{\sigma}_j^R$  de la  $j^{\text{ème}}$  donnée manquante.

L'estimation optimale, selon Fraser de la  $j^{\text{ème}}$  donnée manquante et de sa variance s'écrivent alors comme combinaison linéaire de prévisions, aller et retour :

$$\hat{z}_j = \frac{\hat{z}_j^A (\hat{\sigma}_j^R)^2 + \hat{z}_j^R (\hat{\sigma}_j^A)^2}{(\hat{\sigma}_j^A)^2 + (\hat{\sigma}_j^R)^2} \quad (4.19)$$

$$\hat{\sigma}_j^2 = \frac{(\hat{\sigma}_j^A)^2 (\hat{\sigma}_j^R)^2}{(\hat{\sigma}_j^A)^2 + (\hat{\sigma}_j^R)^2} \quad (4.20)$$

#### 4.3 Application 1 : Cas d'un hydrogramme et d'un pollutogramme synthétiques

Avant d'utiliser toutes les techniques de reconstitution de données décrites précédemment, il nous a semblé intéressant de les appliquer sur des données générées de façon synthétique. L'intérêt ici, est de valider notre méthodologie sur une série de valeurs dont nous connaissons toutes les caractéristiques. Il s'agit d'un hydrogramme et d'un pollutogramme qui obéissent aux conditions normales de continuité et de dérivabilité. Pour les rendre plus proches de la réalité, nous avons rajouté, à chacun d'eux, un bruit différent de distribution normale et d'écart type constant. Nous avons ensuite modifié, volontairement et de façon importante, un certain nombre de valeurs de ces deux courbes synthétiques.

#### 4.3.1 Description des données synthétiques (voir la section 3.3.2)

#### 4.3.2 Critères de performances

##### *Indice et coefficient de performance*

Pour apprécier la qualité des techniques de reconstitution, on utilise un coefficient de performance qui compare la technique ARMA couplée à celle de Fraser, avec la technique utilisant une simple interpolation linéaire :

$$\eta = 1 - \frac{\varepsilon_{ARMA}^2}{\varepsilon_{LIN}^2} \quad (4.21)$$

$\varepsilon_{est}^2$  est la moyenne des écarts quadratiques entre les données estimées ( $x_{ARMA}$ ) par une technique de reconstitution et les données initiales ( $x$ ) non bruitées.  $\varepsilon_{LIN}^2$  est la moyenne des écarts quadratiques entre les données estimées par l'interpolation linéaire ( $x_{LIN}$ ) et les données idéales :

$$\begin{cases} \varepsilon_{ARMA}^2 = \frac{1}{N_{mq}} \sum_{j=1}^N (x_{ARMA}(j) - x(j))^2 \\ \varepsilon_{LIN}^2 = \frac{1}{N_{mq}} \sum_{j=1}^N (x_{mq}(j) - x(j))^2 \end{cases} \quad (4.22)$$

On voit donc que la technique de reconstitution est efficace si l'indice  $\eta$  est positif. Elle devient performante lorsque  $\eta$  approche un.

Pour comparer les deux techniques entre elles, on introduit le paramètre P de performance :

$$P = \frac{\varepsilon_{LIN}^2 - \varepsilon_{ARMA}^2}{\varepsilon_{LIN}^2} \times 100 \quad (4.23)$$

### 4.3.3 Résultats

Afin de tester la méthodologie de reconstitution, on a volontairement simulé sept données manquantes parmi les quarante données de concentration de MES qu'on avait préalablement bruitées : Deux données manquantes situées sur chacune des pointes, une sur la partie descendante du premier pic puis deux, vers la fin de l'événement, c'est à dire la zone qui peut représenter la concentration de polluants en temps sec.

Quoique la reconstitution par interpolation linéaire ici soit acceptable (figure 4.2b), il est important de constater que la reconstitution par ARIMA couplé à celle de Fraser (figure 4.2a) est plus performante. En effet, un calcul sur les sept valeurs reconstituées donné une performance de cette dernière méthode de l'ordre de 31% par rapport à celle d'interpolation linéaire. Par ailleurs, la première pointe du pollutogramme a été parfaitement reconstituée par la méthode ARIMA. Par contre, la reconstitution par interpolation linéaire ne peut que sous estimer, par construction, toute pointe d'une courbe. C'est d'ailleurs ce qui a été remarqué pour les deux pointes du pollutogramme. Une surestimation de la seconde pointe (figure 4.2a) est observée dans le cas de la reconstitution par ARIMA, néanmoins il ne figure aucun décalage dans le temps.

Pour les données manquantes situées en dehors des pointes, on peut affirmer que les deux méthodes de reconstitution se comportent de la même manière et qu'il est suffisant de faire une reconstitution de ces types de données, par simple interpolation linéaire.

#### 4.4 Application 2 : Cas d'un pollutogramme réel

##### 4.4.1 Description du bassin urbain (voir 3.4.1)

##### 4.4.2 Critères de performances

Dans le cas des mesures réelles, ces critères sont moins évidents à définir car on ne connaît pas a priori les données «vraies». Néanmoins, deux types de critères peuvent être définis.

Pour apprécier la qualité des techniques de reconstitution, on utilise un coefficient de performance qui compare la technique ARMA couplée à celle de Fraser, avec la technique utilisant une simple interpolation linéaire :

$$\eta = 1 - \frac{\varepsilon_{ARMA}^2}{\varepsilon_{LIN}^2} \quad (4.24)$$

$\varepsilon_{est}^2$  est la moyenne des écarts quadratiques entre les données estimées ( $x_{ARMA}$ ) par une technique de reconstitution et les données réelles filtrée ( $x$ ) non bruitées.  $\varepsilon_{LIN}^2$  est la moyenne des écarts quadratiques entre les données estimées par l'interpolation linéaire ( $x_{LIN}$ ) et les données réelles filtrées :

$$\begin{cases} \varepsilon_{ARMA}^2 = \frac{1}{N_{mq}} \sum_{j=1}^N (x_{ARMA}(j) - x(j))^2 \\ \varepsilon_{LIN}^2 = \frac{1}{N_{mq}} \sum_{j=1}^N (x_{mq}(j) - x(j))^2 \end{cases} \quad (4.25)$$

On voit donc que la technique de reconstitution est efficace si l'indice  $\eta$  est positif. Elle devient performante lorsque  $\eta$  approche un.



Pour comparer les deux techniques entre elles, on introduit le paramètre P de performance :

$$P = \frac{\varepsilon_{LIN}^2 - \varepsilon_{ARMA}^2}{\varepsilon_{LIN}^2} \times 100 \quad (4.26)$$

P exprime donc l'amélioration apportée par ARMA, comparativement à l'interpolation linéaire.

#### 4.4.3 Résultats

On applique ici une méthodologie de reconstitution à la série de valeurs de concentration de MES mesurées à Rhéaume. Pour rendre compte de sa performance, on a ainsi simulé 11 données manquantes. Deux vers 5h, quatre autour de 6h, au niveau de la pointe du pollutogramme, deux vers 6h30', en période de récession et quatre à 7h30', au niveau de la seconde pointe du pollutogramme.

Comme on le voit, la reconstitution utilisant la méthodologie ARMA couplée à celle de Fraser (figure 4.3a) s'est beaucoup mieux comportée que celle utilisant l'interpolation linéaire (figure 4.3b). Cette dernière ayant l'inconvénient de sous estimer presque systématiquement les pointes.

#### 4.5 Conclusion

Les résultats ont été comparés à l'aide de critères de performance quantitatifs ainsi que qualitatifs. La méthodologie poursuivie donne de bons résultats aussi bien pour la série synthétique bruitée que pour la série réelle. Pour la série réelle le procédé a permis d'améliorer nettement les coefficients d'autocorrélation.

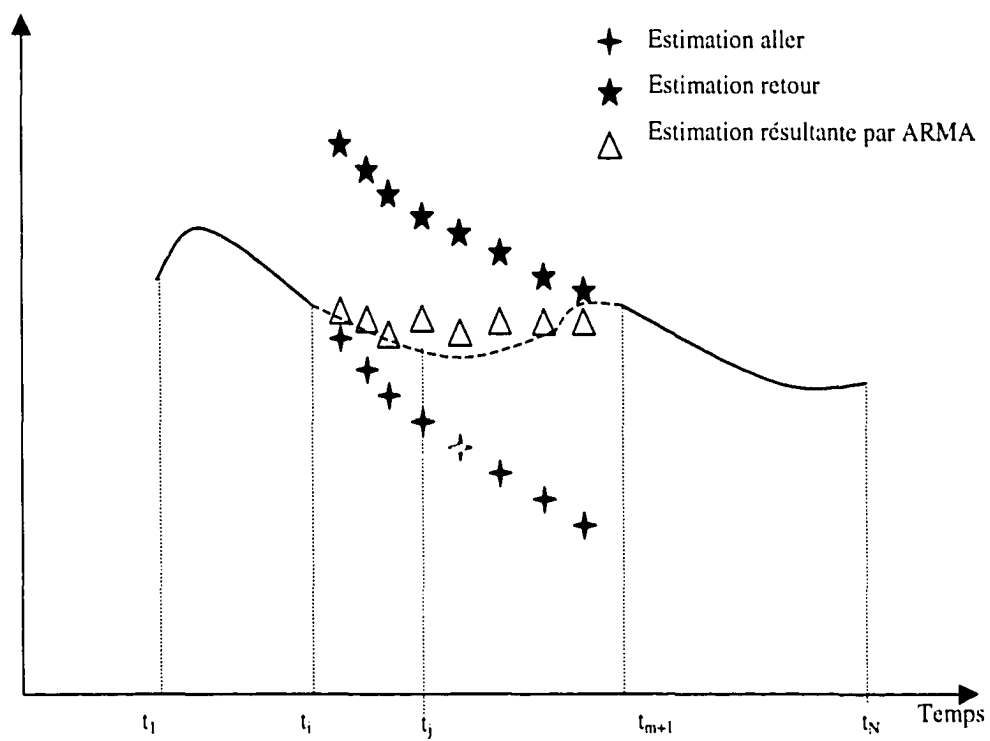


Figure 4.1 Méthode de reconstitution de données par ARMA/Fraser

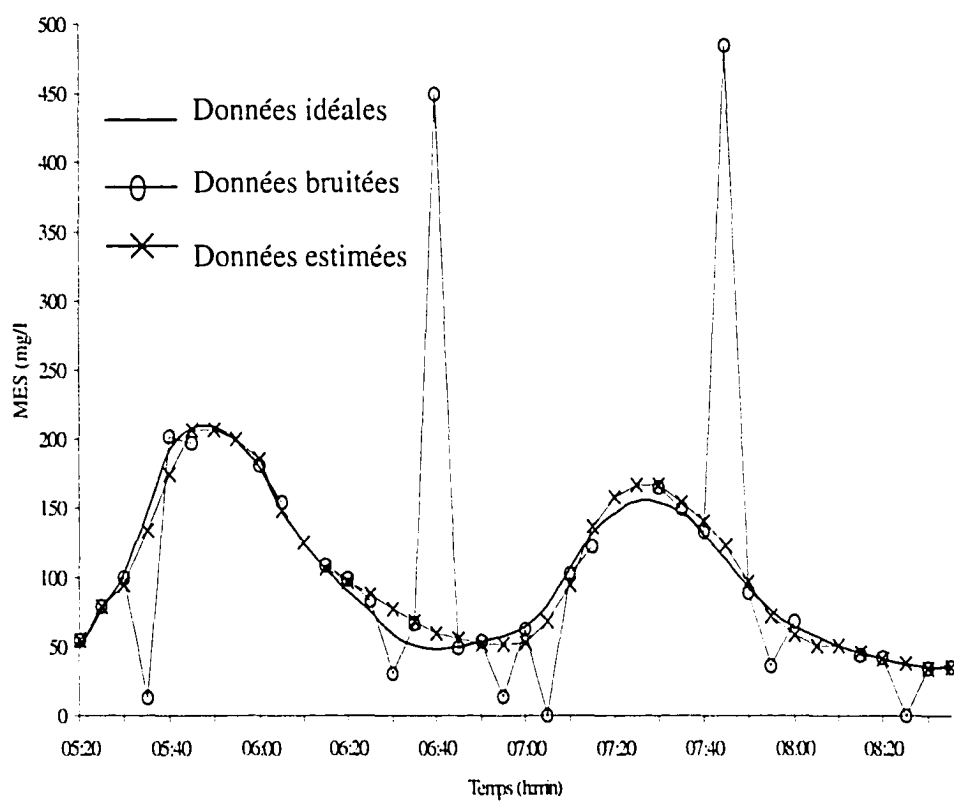


Figure 4.2a Reconstitution des données manquantes simulées pour le  
pollutogramme synthétique(ARIMA)

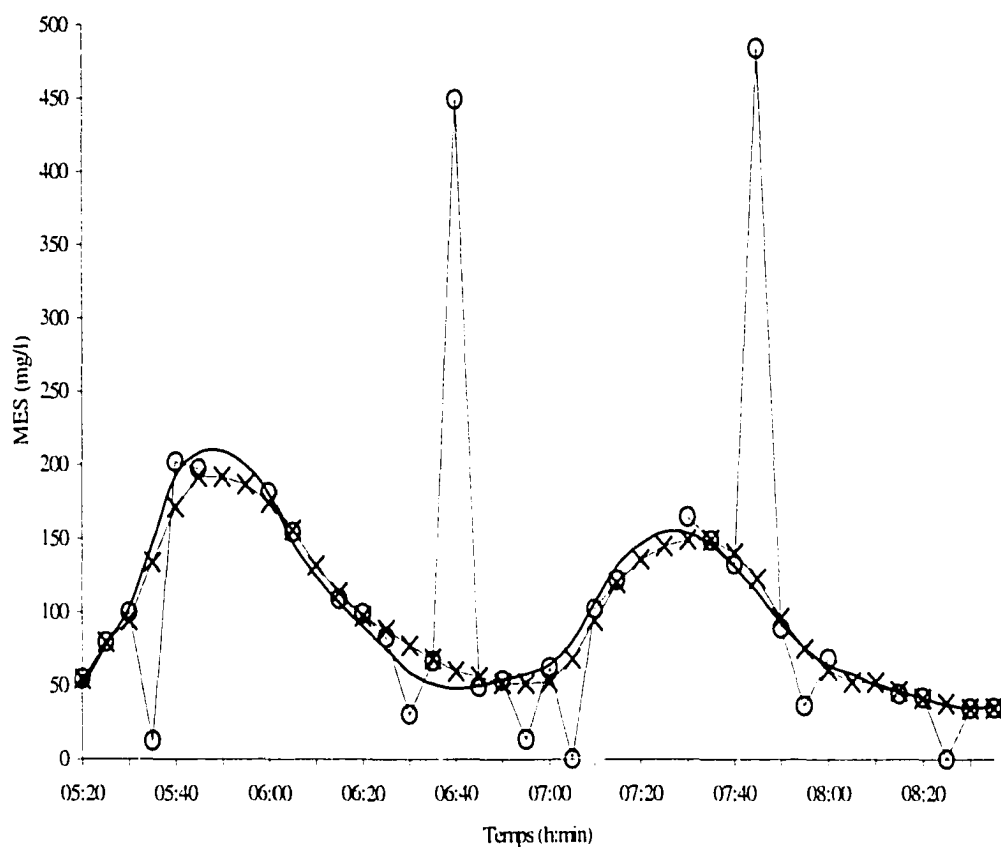


Figure 4.2b Reconstitution des données manquantes simulées pour le pollutogramme synthétique(interpolation linéaire)

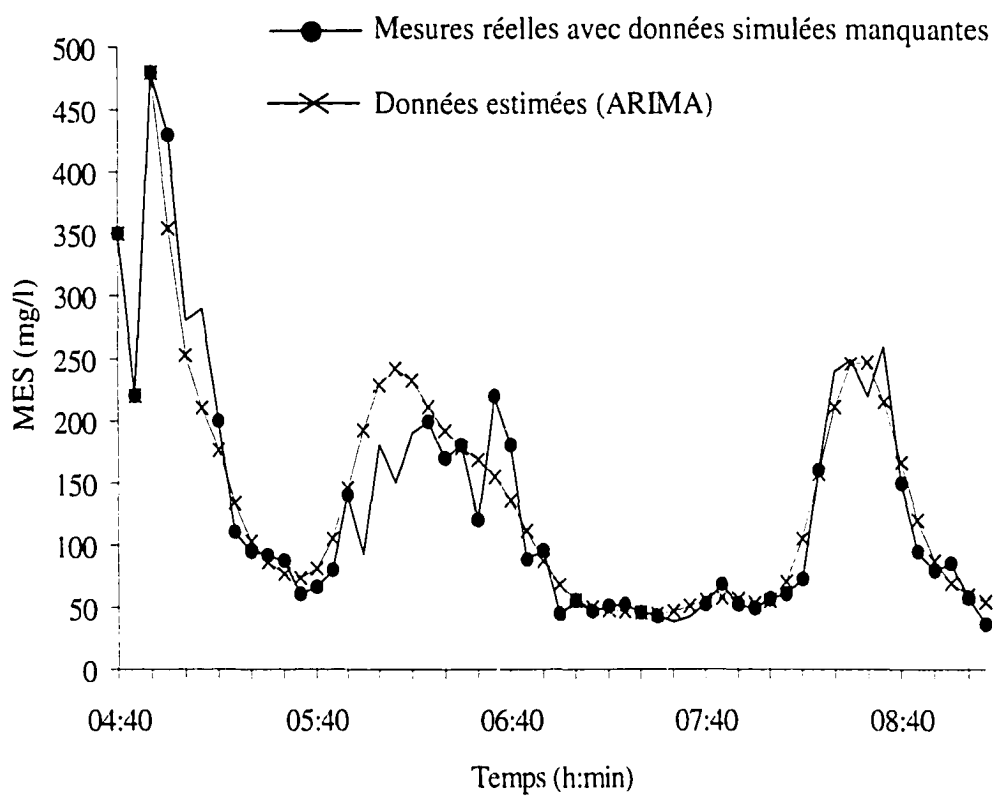


Figure 4.3a Reconstitution des données manquantes. Cas d'un pollutogramme réel

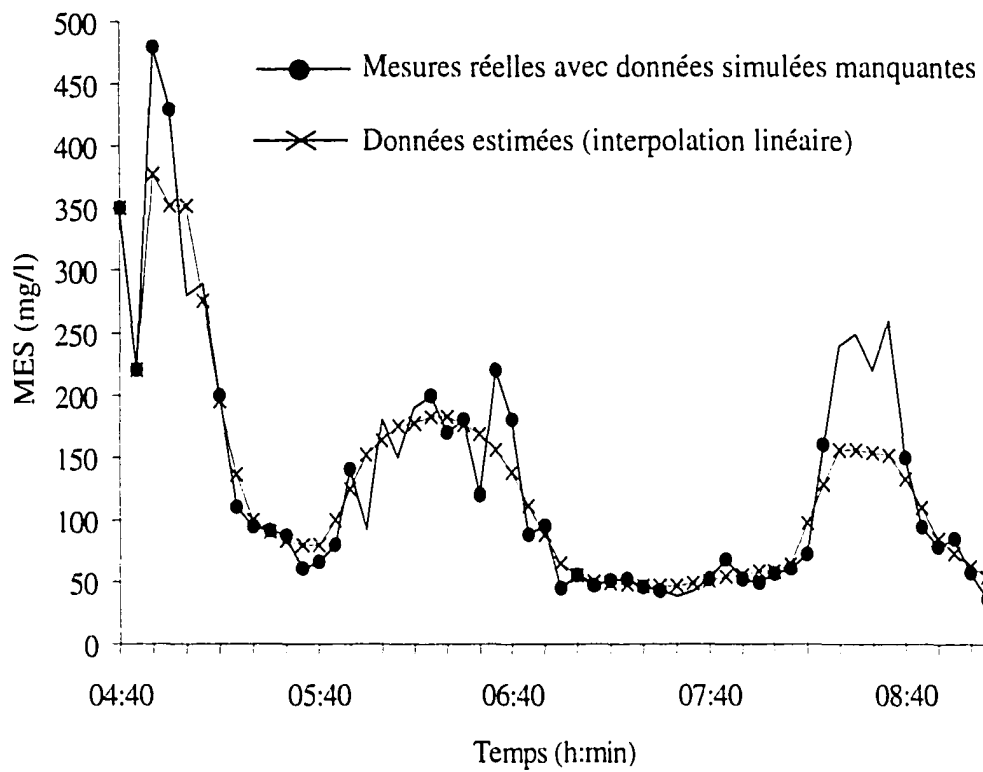


Figure 4.3b Reconstitution des données manquantes. Cas d'un pollutogramme réel

## CONCLUSION ET RECOMMANDATIONS

Ce travail nous a permis de trouver une combinaison de filtres qui donne les meilleurs résultats pour la validation des données que l'on pourrait par la suite appliquer pour la gestion en temps réel.

Les modèles utilisés pour valider les données sont des filtres univariés qui ont été appliqués à des séries chronologiques de concentration en MES avec un pas de temps de cinq minutes. Les résultats ont ensuite été comparés à l'aide de critères qualitatifs et quantitatifs.

Les combinaisons de filtres présentées donnent de bons résultats. Les deux meilleures combinaisons sont : « WIN5-2 »+« FRAS »+ « FRÉQ » et « WIN7-4 »+« FRAS »+ « FRÉQ ».

Ces combinaisons consistent en un pré-lissage par un filtre ébarbeur Winsoriseur puis une application par la suite du filtre de Fraser qui élimine les valeurs aberrantes puis finalement le filtre fréquentiel, qui améliore le lissage. Nous avons expérimenté sur une série synthétique bruitée pour essayer de retrouver le signal lisse de départ, ensuite nous les avons appliqués sur une série réelle. Nous avons améliorés nettement les coefficients d'autocorrélation.

Une technique performante est utilisée pour obtenir la meilleure estimation des données manquantes successives par pondération des valeurs obtenues par deux modèles autorégressifs. Le premier opère dans le sens croissant du temps l'autre dans le sens décroissant. La technique MISP qui utilise successivement deux filtres de Kalman, interactifs entre eux, un pour filtrer la variable d'état l'autre pour identifier les paramètres du modèle, est appliquée dans les deux sens «Aller » et «Retour ».



## BIBLIOGRAPHIE

ABRAHAM, B., et LEDOLTER, J. (1983). *Statistical Methods for Forecasting*. John Wiley & Son, Inc, New York.

AKAIKE, H., (1974). A new look at the Statistical Model Identification. *Automatic Control*. Volume 19. N° 6. Pages 716-723

BARNETT, V., et LEWIS, T. (1984). *Outliers in Statistical Data*, 2<sup>e</sup> édition. John Wiley & Son, Inc, New York.

BENNIS S. (1987). *Prévision des Crues par l'Analyse des Séries Chronologiques*. Thèse de Ph.D., Université de Sherbrooke, Sherbrooke. pp. 5-55.

BENNIS S., et BRUNEAU. (1993a). *Comparaison des méthodes d'estimation des débits journaliers*. *Canadian Journal of Civil Engineering* Montréal. VOL 20 p 480-489.

BENNIS S., et BRUNEAU. (1993b). *Amélioration des méthodes d'estimation des débits journaliers*. *Canadian Journal of Civil Engineering* Montréal. VOL 20. p 490-499.

BENNIS S., et NARUT K. (2000). *Multivariate Technique for Validating Historical Hydrometric Data with Redundant Measurements*. *Nordic Hydrology*. Montréal.

BERRADA F. (1998). *Contribution à l'Analyse des Séries Chronologiques en Hydrologie et Élaboration de Procédés de Validation et Reconstitution des Données Hydrométriques*. Doctorat d'état Es Sciences Physiques, Université Hassan II, Faculté des Sciences Ain Chok, Casablanca, Maroc.

BERRADA, F., BENNIS, S., et GAGNON, L. (1996). *Validation des Données Hydrométriques par des Techniques Univariées de Filtrage*. *Canadian Journal of Civil Engineering* Montréal. p 872-892.

BOX, G. E. P., et JENKINS, G. M. (1976). *Time Series Analysis : Forecasting and Control*. Édition révisée. Holden-Day, Oakland. pp. 46-82.

- CHOCAT, (1997). *Encyclopédie de l'Hydrologie urbaine et de l'assainissement*. ISBN : 2-7430-0126-7. P. 183.
- DAVIES, (1983). A recursive approach to prony parameter estimation. *Journal of sound and vibration*, 89 :571-583
- DRUEN, M. ET DOOGE, J.C.I., (1984). An effective and Robust Method for Estimating Unit Hydrograph Ordinates. *Journal of Hydrology*. N° 70. Pages 1-24.
- EDWARD L. THACKSTON, FELLOW, ASCE, et ANDREW MURR, ASSOCIATE MEMBER, ASCE. CSO Control Project Modifications Based on Water Quality Studies. *Journal of environmental engineering/* (October 1999), pp 979-987.
- FITCH, M.A., et McBEAN, E.A. (1991). Multi-day flow forecasting using the Kalman filter. *Revue canadienne de génie civil*, 18 : 320-327.
- GAGNON, I., (1994). *Filtrage des Séries Hydrométéorologiques*. Mémoire de Maîtrise en Technologie des Systèmes. ETS, Montréal.
- GILLES RIVARD, 1998. *Concepts et applications*. Alias Communication Design Inc. Montréal (Québec).
- HUBER, W.C., et DICKINSON, R.E., (1988). *Storm Water Management Model : User's manual*. Version 4 University of Florida, Gainesville, Florida.
- JEAN-PIERRE.P , (1984) : *Modélisation de la pollution des eaux de ruissellement pluvial urbain : étude bibliographique*. Document du laboratoire régional de l'Ouest Parisien, 1984.
- JEAN-PIERRE.P; JEAN.R, (1987) : *Pollution des eaux de ruissellement urbain en zone urbaine*. Laboratoire central des ponts et chaussées, Paris, 1987.
- JEFFREY D.SHARON CHAIRMAN, (1989) *Task Force On Cso Pollution Abatement. Combined sewer overflow pollution abatement, manuel of practice* NO. FD-17, Water Pollution Control Federation 601 Wythe Street, Alexandria, VA 22314- 1994.
- KALMAN, R. E., et BUCY R. S. (1961). *New Results in Linear Filtering and Prediction Theory*. *Journal of Basic Engineering* (Translations of ASME, Series D), 83 : 95-107.
- KACHROO, R.K. et LIANG, L.G.C., (1992). *Riverflow Forecasting. Part II. Algebraic Development of Linear Modelling Technics*. *Journal of Hydrology*. N° 133. Pages 17-40.
- KOTHANDARAMAN V, (1972) · Water quality characteristics of storm sewer discharges and combined sewers. Illinois state Water survey.

LARRY A. ROESNER AND PAUL TRAINA (1994). *Overview of federal law and USPA regulations for urban runoff*. *Water Science and Technology*, VOL 29, n°. 1-2, PP 445-454.

MCCUEN, R.H., et SNYDER, W.M., (1986). *Hydrologic Modelling : Statistical Methods and Applications*. Prentice-Hall, Englewood Cliffs.

MINISTÈRE DE L'ENVIRONNEMENT DU QUÉBEC, rév. (1992). *Critères de la qualité de l'eau. Service d'évaluation des rejets toxiques et Direction de la qualité des cours d'eau*, Ministère de l'Environnement du Québec, Québec, 423 p.

MINISTÈRE DE L'ENVIRONNEMENT DU QUÉBEC, (1990) (rév. 1992). *Méthodologie de calcul des critères de qualité de l'eau pour les substances toxiques*. Ministère de l'Environnement du Québec, Québec, 147 p.

MINISTÈRE DE L'ENVIRONNEMENT DU QUÉBEC, 1991 (rév. 1996). *Méthode de calcul des objectifs environnementaux de rejet pour les contaminants du milieu aquatique*. Ministère de l'Environnement du Québec, Québec, 21 p.

MINISTÈRE DE L'ENVIRONNEMENT DU QUÉBEC, (1993 préliminaire, rév. 1996). *Guide pour l'évaluation et la réduction des toxiques*. Ministère de l'Environnement du Québec, Québec, 72 p.

NGUYEN V.T.V. et BISSON J.L., (1998). *Validation en Temps Réel des Données des Apports Naturels Journaliers pour la Gestion des Réservoirs*. *Journal Canadien du génie Civil*. No 25, 1096- 1102.

O'CONNELL, P.E. (Éditeur), (1980). *Real-Time Hydrological Forecasting on Large Catchments*. N° 65. Pages 125-137.

PANKRATZ, A. (1983). *Forecasting with univariate Box-Jenkins models. Concepts and cases*. John Wiley & Son Inc., New York, Chichister, Brisbane, Toronto et Singapore. pp. 224-230.

PERRAULT, L., ROY, R., MATHIER, L., et BOBÉE, B., (1995). *La Combinaison de Modèles Appliqués à la Validation en Temps Réel des Apports Naturels aux Réservoirs Hydriques*. *Revue canadienne de Génie Civil*, Vol. 22 No. 5, 934-944.

PIATYSZEK, E., (1998). *Détection de Dysfonctionnements en Système Hydrographique : application aux réseaux d'assainissement*. Thèse de doctorat. ENSMSE, No. 187ID, 410 p.

RADIX, C. (1984). *Filtrage et Lissage Statistiques Optimaux Linéaires*. Cepadues éditions, Toulouse. pp. 46-73 et 163-165.

RAGOT, J., DAROUACH, M., MAQUI, D., et BLOCH, G. (1990). *Validation des Données et Diagnostic*. Édition Hermès, Paris. pp. 79-92.

ROY, R., GAUTHIER, R., et HÉRU, M. (1992). *Hydrometeorological Data Quality Improvement by Implementing an Autonomous Acquisition Network in Barren Areas. Hydraulic engineering software IV. Computer Techniques and Applications*, Southampton et Boston. pp. 351-357.

SWARTZ, C.L.E. (1989). *Data Reconciliation for Generalized Flowsheet Applications*. American Chemical Society, national meeting, Dallas, 10 au 14 avril.

SOLIMANN, S.S., et SRINATH, M.D. (1990). *Continuous and Discrete Signals and Systems*. Prentice Hall, Englewood Cliffs.

SCHILLING et MARTENS, 1986. Recursive state and parameter estimation with applications in water resources. *Applied mathematical modelling*, 10 :433-437.

TODINI, E. (1978). *Mutually Interactive State-Parameter (MISP) Estimation in Application of Kalman Filter to Hydrology, Hydraulics and Water Resources*. Éditeur : C. L. Chiu. *Proceedings of the American Geophysics Union (A G U) Chapman Conference*, University of Pittsburgh. pp. 135-151.

URBAN WET WEATHER POLLUTION (1996), *Controlling sewer overflows and stormwater runoff*, Water Pollution Control Federation 601 Wythe Street, Alexandria, Virginia 22314- 1994.

VAN DEN ENDEN, A. W. M., et VERHOECKX, N. A. M. (1992). *Traitement Numérique du Signal*. Une Introduction. Paris, Milan, Barcelone et Bonn. pp. 222-231.

WINKLER, R.L., et MAKRIDAFIS, S. (1983). *The combination of forecasts*. *Journal of Royal Statistic Society A*, 46 : 150-157.

WALD, A., (1947). *Sequential Analysis*. John Wiley Sons. New-York.

YOUNG, P.C., (1970). *An Instrumental Variable Method for Real-Time Identification of Noisy Process*. *Automatika*, N° 6. Pages 271-287.

YOUNG, P.C., (1974). *Applying parameter estimation to dynamic systems. Control engineering*. N° 10. Pages 119-124.